Evolution of Conditional Cooperation

Madgwick, Philip

Award date:
2019

Awarding institution:
University of Bath

Evolution of conditional cooperation

Philip Graham Madgwick

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Milner Centre for Evolution, Department of Biology & Biochemistry

September 2019

DECLARATION OF ANY PREVIOUS SUBMISSION OF THE WORK

The material presented here for examination for the award of a higher degree by research has not been incorporated into a submission for another degree.

DECLARATION OF AUTHORSHIP

I am the author of this thesis, and the work described therein was carried out by myself personally, with the exception of the Papers where the work was carried out in collaboration with other researchers; the details of which are included in the ‘Statement of Authorship’ prefacing each Paper.

Acknowledgements

I would to extend my thanks to all my friends, family and colleagues for their indispensable support over the last four years. In particular, for the completion of this thesis, I would like to thank Elizabeth Milne, Laurence Belcher and Jason Wolf for comments/proof-reading, and more generally to Laurence Belcher for his comradeship. I would like to extend a special thanks to Jason Wolf, my supervisor, for his unwavering patience and confidence as my research project morphed beyond all recognition of its premise, and for giving me free rein to follow my own curiosity.

Abstract

Cooperation presents evolutionary theory with an interesting puzzle: why should an actor help a recipient? Evolutionary explanations of cooperation rely on finding ways through which cooperative behaviour benefits the actor, by considering the selection on a gene for cooperation. A critical variable is the relatedness between the actor and the recipient, which determines the probability that they share the gene for cooperation. Consequently, a classic expectation is that an actor should only cooperate on the condition that the relatedness between the actor and the recipient ensures a net gain for the gene for cooperation. However, such conditional cooperation has been dismissed as an unimportant part of the evolutionary theory of cooperation because the simpler explanation of unconditional cooperation is usually sufficient. The scepticism of the relevance of conditional cooperation is often based on two premises: first there is very little evidence of its importance, and second conditional cooperation lacks a plausible mechanism for its operation and persistence. In this thesis by publication, I provide a rebuttal using experimental evidence of conditional cooperation in the social amoeba *Dictyostelium discoideum*, developing theoretical models to examine the possible mechanisms for conditional cooperation and comparatively situating these arguments within the broader context of other study systems. Using five papers (and a comment on each), I present the case that conditional cooperation should not be so easily dismissed, as it seems likely to be a more important part of the evolutionary explanation of cooperation in nature than current evidence might suggest.

Table of Contents

Introduction

Cooperative behaviour is widespread throughout the natural world, and yet it presents evolutionary theory with a special difficulty (Hamilton, 1964a; b; Williams, 1966; Leigh, 1977; Alexander & Borgia, 1978; Bull & Rice, 1991; Frank, 2003; Sachs *et al.*, 2004; West *et al.*, 2007a; Bourke, 2011). In cooperation, an actor provides a benefit to a recipient, so why does the actor do it? Evolutionary explanations of cooperation rely upon finding a way in which the actor gains from cooperation (West *et al.*, 2007a). These explanations can be broadly divided into two, which can be understood from a genetic perspective using Hamilton's rule, by considering the balance of selection on a gene for cooperation (Hamilton, 1963; Charnov, 1977; Grafen, 1985). First, the actor may indirectly gain when the cooperative behaviour is altruistic (Hamilton, 1964a), such that the actor pays a direct fitness cost ($c < 0$) to provide a fitness benefit to the recipient ($b > 0$). The gene for cooperation may be favoured by selection despite paying a direct fitness cost in the actor because the same gene is also carried by the recipient due to relatedness between the actor and recipient (r). The most prevalent source of relatedness in nature is common ancestry, and so relatedness can be described with a kinship coefficient (Haldane, 1955; Williams & Williams, 1957; Hamilton, 1963). However, in genetic terms, relatedness can be thought of more generally as the probability that the actor and recipient share the gene over and above the population average (and so relatedness can be positive or negative, which the kinship definition can overlook) (Grafen, 1985). Consequently, the gene for cooperation reaps an indirect fitness benefit from the recipient is accordance with the benefit-weighted probability of sharing the gene (rb). In this way, the gene for cooperation is favoured when it has positive inclusive fitness, which is calculated as the addition of direct and indirect fitness: $rb - c > 0$. When relatedness is negative, the gene for cooperation is never favoured (because $rb < 0$ making inclusive fitness always less than zero). Second, the actor may directly gain when the cooperative behaviour is mutually beneficial to both the actor and the recipient (Hamilton, 1964a). In this case, the actor

reaps a direct fitness benefit (or negative cost; $c < 0$) and so always has positive inclusive fitness when relatedness is positive (or zero; because both $rb > 0$ and $c > 0$). However, when relatedness is negative, the direct fitness negative costs ($c < 0$) must outweigh the indirect fitness positive benefits ($b > 0$, given $r < 0$) for cooperation to be favoured (*i.e.* for inclusive fitness to be positive). Therefore, for both altruistic and mutually beneficial behaviour, relatedness always has the potential to reverse the direction of selection on a gene for cooperation, making Hamilton's rule a useful approach.

Evolutionary explanations of cooperative behaviour have given particular focus to altruism (Williams, 1966; Hamilton, 1972; Wilson, 1975b; Frank, 2003; West *et al.*, 2007a; Bourke, 2011). Many of the most eye-catching forms of cooperation in nature are interpreted as altruism, including sterile castes in social insects (Hamilton, 1964b), helping at the nest in birds (Skutch, 1935) and communal care in mammals (Riedman, 1982). Further, altruism appears more difficult for evolutionary theory to explain because of the problem of cheaters (Foster, 2004; Sachs *et al.*, 2004; Travisano & Velicer, 2004; West *et al.*, 2006, 2007a; Bourke, 2011; Ghoul *et al.*, 2014). Altruistic behaviour involves an actor paying a cost to give a recipient a benefit, and so altruistic behaviour is vulnerable to a cheater that is the recipient of the benefit without ever paying its costs (Hamilton, 1963). This focus may in part be because the threat of cheaters can be easily studied both theoretically and experimentally. Theoretically, cheaters can be studied using simple models like a prisoner's dilemma (Hamilton, 1971; Maynard Smith & Price, 1973; Maynard Smith, 1979; Axelrod & Hamilton, 1981), which show that selection does not tend to favour cheaters when relatedness is sufficiently high (Travisano & Velicer, 2004; West *et al.*, 2006; Bourke, 2011). Experimentally, cheaters can be studied through competition experiments between a wildtype and mutant individual/strain (which can easily be generated, *e.g.* by knock-out out a cooperative phenotype through random mutagenesis; *e.g.* Velicer *et al.*, 2000; Foster *et al.*, 2004; Griffin *et al.*, 2004). These studies show that a gene for altruism can be vulnerable to

invasion by a gene for cheating, which can prevent individuals from reaping the potential benefits of altruistic behaviour.

In recent years, there has been some confusion surrounding the focus of evolutionary explanations of cooperation on the problem of altruistic behaviour due to the conceptual conflation of cooperation and altruism. Part of this confusion stems from the forms of cooperation that appear altruistic at first inspection, but have been revealed to be mutually beneficial after detailed study, like sentinel behaviour in meerkats (Clutton-Brock *et al.*, 1999), cooperative breeding in wasps (Queller *et al.*, 2000) and worker policing in some social insects (Wenseleers & Ratnieks, 2006). Symptomatic of this confusion is the continuing description of these systems as if they were vulnerable to cheating, which is not the case for mutually beneficial behaviour. A gene for cheating would be unable to invade a gene for mutual benefit because, whilst it would gain the indirect fitness benefit of being a recipient, it would not gain the direct fitness benefit (or negative cost) of being the actor, and so have lower inclusive fitness. But another part of this confusion also stems from the over-application of the concept of altruism to biological systems because of the fashionable way that it can construe cooperation as paradoxically vulnerable to cheaters, leading to confusing terminology like reciprocal altruism, weak altruism and altruistic punishment, which are all forms of mutual benefit (West *et al.*, 2007b).

When considering the evolution of a particular cooperative behaviour, the confusion surrounding its vulnerability to cheating is exacerbated when considering 'conditional cooperation' (*sensu* Marshall, 2015) where the behaviour is varied in different situations (Maynard Smith, 1976; Axelrod & Hamilton, 1981; Axelrod, 1984; Axelrod *et al.*, 2004; Doebeli *et al.*, 2004; Doebeli & Hauert, 2005). The most-studied form of conditional cooperation is facultative cooperation, where a gene can switch behaviour between cooperating or cheating with social partners (Velicer *et al.*, 2000; Abbot *et al.*, 2001; Griffin *et al.*, 2004; Santorelli *et al.*, 2008; Khare & Shaulsky, 2010; Ferrari *et al.*, 2015; Pollak *et al.*, 2016; Nair *et al.*, 2018). But conditional cooperation need not be all-or-none, as a gene could

also quantitatively adjust the level of cooperation from anywhere between what appears to be altruism or cheating dependent upon the social situation (Madgwick *et al.*, 2018). The vulnerability of conditional cooperation to cheating is, arguably, reduced because conditionality permits the cooperation to persist in specific situations where cheating may generally be favoured in the average social situation. But the focus on cheating is really missing the key question about balance of selection on a gene, where it is anticipated that a gene for cheating would be outcompeted by a gene for conditional cooperation (which acts as a cheater in some situations and as a cooperator in others). In this way, conditional cooperation does not avoid cheating, but instead permits an individual to better exploit their social situation come what may.

Conditional cooperation has received some theoretical and experimental study, but the empirical tests of theory have focused on simple qualitative predictions (Reeve & Dugatkin, 1998; McNamara, 2013). Consequently, the only forms of conditional cooperation where theory has been used to derive specific predictions about the patterns of conditional cooperation involve facultative cooperation comparing two social situations (which are usually high and low relatedness between social partners; see above) (Velicer *et al.*, 2000; Abbot *et al.*, 2001; Griffin *et al.*, 2004; Santorelli *et al.*, 2008; Khare & Shaulsky, 2010; Ferrari *et al.*, 2015; Pollak *et al.*, 2016; Nair *et al.*, 2018) rather than testing quantitative predictions. In general, there has been a scepticism that conditional cooperation is really important for the evolutionary explanation of cooperation in biological systems because 'recognition' of social partners (or other features of the social situation) is an unnecessarily complicated hypothesis for many systems where there is high relatedness (*e.g.* instead relying on limited dispersal) (Grafen, 1990; Rousset & Roze, 2007; West *et al.*, 2007a, 2011; Cornwallis *et al.*, 2009; West & Gardner, 2013). There are several factors that are often used to support this scepticism. First, there is little evidence that conditional cooperation is important. In counter, I would argue that this is a premature assessment because many empirical studies simply assume that cooperation is unconditional at some fixed level (West *et al.*, 2007a; Foster, 2009; Bourke,

2011). Further, there is likely to be a publication bias; the patterns of conditional cooperation are, arguably, harder to study than the threat of obligate cheating because conditional cooperation must be studied within natural (or naturalised laboratory) settings to avoid biasing the social signals/cues that individuals are using to adjust their level of cooperation, and cannot simply rely on a simple competition experiment between a wildtype and mutant (knock-out) strain. Second, it is unclear how conditional cooperation would work mechanistically – both in the functional and evolutionary senses. Whilst this scepticism is reasonable for some classic study systems in animals, molecular biology has revealed how signal-receptor proteins can permit interacting individuals to characterise their social situation (Haig, 1996; Springer *et al.*, 2011). However, the genes that we know about that encode for signal-receptor proteins that modulates a cooperative behaviour are typically suggested to be greenbeards (Gardner & West, 2010; Madgwick *et al.*, 2019), which are often thought to only permit simplistic social traits rather than the complicated traits that we observe (West & Gardner, 2013). Further, any recognition genes (*i.e.* greenbeard or kin-selected) are anticipated to be subject to Crozier's paradox, whereby more common types are favoured, leading one signal-receptor variant to go to fixation, which prevents conditional cooperation from persisting (Crozier, 1986; Rousset & Roze, 2007).

This thesis by publication contributes to the evolutionary explanation of cooperation by addressing the puzzle of conditional cooperation. I start from the observation that different strains of *Dictyostelium discoideum* appear to exhibit conditional cooperation across frequencies (Buttery *et al.*, 2009). Paper 1 presents a game theoretic model that describes the patterns of conditional cooperation that are expected to evolve under natural selection, and tests these patterns as predictions to demonstrate a close fit of theory and data within chimeric mixes of two strains. Comment 1 explores how these results tie into the logic of simple bimatrix games of the prisoner's dilemma, snowdrift and stag hunt. Paper 2 tests predictions of the patterns of conditional cooperation in chimeric mixes of *D. discoideum* when there are more than two strains, showing how nonadaptive constraints on the flexibility of conditional

cooperation can be beneficial. Comment 2 uses these findings to suggest the likely mechanism by which strains are adjusting their level of cooperation. Paper 3 generalises the game theoretic model for *D. discoideum* to explain the logic of conditional cooperation across different systems, suggesting how conditional cooperation can lead to negative frequency-dependent fitness which can overcome Crozier's paradox. Comment 3 explores how the patterns and explanation of conditional cooperation are robust to real-world considerations. Paper 4 examines the experimental evidence of greenbeard genes across study systems, which can be critical in the mechanism of conditional cooperation. Comment 4 makes further mathematical remarks on greenbeard theory, showing how different conceptions of a greenbeard gene suggest different properties. Paper 5 takes inspiration from greenbeards to consider how to improve the design of gene drives, as an application of the logic of conditional cooperation to a practically useful end. Starting from the greenbeard concept, Comment 5 develops the concept of an unbeatable gene as a design objective for gene drives. In an overarching discussion, I bring together my findings. Therefore, starting from a puzzling observation, this thesis presents the case that conditional cooperation is likely to be a more important part of the evolutionary explanation of cooperation in nature than current evidence might suggest.

Paper 1

This declaration concerns the article entitled:			
Strategic investment explains patterns of cooperation and cheating in a microbe			
Publication status (tick one) **Draft manuscript** ☐ **Submitted** ☐ **In review** ☐ **Accepted** ☐ **Published** ✓			
Publication details (reference)	**Madgwick, P. G.**, L. J. Belcher, B. Stewart, C. R. L. Thompson, and J. B. Wolf. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. Proceedings of the National Academy of Sciences 115:E4823–E4832.		
Copyright status (tick the appropriate statement) I hold the copyright for this material ✓ Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Formulation of ideas: 33%: PGM contributed to the research concept. Design of methodology: Experimental 0% Theoretical 33%: PGM contributed to the model design, significantly contributed to model analysis and contributed to the interpretation of data. Experimental work: 0% Presentation of data in journal format: 40%: PGM significantly contributed to drafting, revising and editing the manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	PGMadgwick	**Date**	07/11/2019

Data access statement: email jbw22@bath.ac.uk

Strategic investment explains patterns of cooperation and cheating in a microbe

Authors: Philip G. Madgwick[1]†, Balint Stewart[2]†, Laurence J. Belcher[1], Christopher R.L. Thompson[2*], and Jason B. Wolf[1*]

Affiliations:

[1] Milner Centre for Evolution and Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

[2] Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK

[*] Correspondence to: jason@evolutionarygenetics.org and christopher.thompson@ucl.ac.uk

† These authors contributed equally to this work

Abstract

Contributing to cooperation is typically costly, while its rewards are often available to all members of a social group. So why should individuals be willing to pay these costs, especially if they could cheat by exploiting the investments of others? Kin selection theory broadly predicts that individuals should invest more into cooperation if their relatedness to group members is high (assuming they can discriminate kin from non-kin). To better understand how relatedness affects cooperation, we derived the 'Collective Investment' game, which provides quantitative predictions for patterns of strategic investment depending on the level of relatedness. We then tested these predictions by experimentally manipulating relatedness (genotype frequencies) in mixed cooperative aggregations of the social amoeba *Dictyostelium discoideum*, which builds a stalk to facilitate spore dispersal. Measurements of stalk investment by natural strains correspond to the predicted patterns of relatedness-dependent strategic investment, wherein investment by a strain increases with its relatedness to the group.

Furthermore, if overall group relatedness is relatively low (*i.e.*, no strain is at high frequency in a group) strains face a scenario akin to the 'Prisoner's Dilemma' and suffer from insufficient collective investment. We find that strains employ relatedness-dependent segregation to avoid these pernicious conditions. These findings demonstrate that simple organisms like *D. discoideum* are not restricted to being 'cheaters' or 'cooperators', but instead measure their relatedness to their group and strategically modulate their investment into cooperation accordingly. Consequently, all individuals will sometimes appear to cooperate and sometimes cheat due to the dynamics of strategic investing.

Keywords: cooperation; conflict; game theory; cheating; kin selection

Significance statement

Contributing to cooperation is costly, while its rewards are often available to all members of a social group. Therefore, cooperation is vulnerable to exploitation by individuals that do not contribute, but nevertheless share the benefits. So why contribute to cooperation? This dilemma can be resolved if individuals modulate their 'investment' into cooperation dependent on whether benefits go to relatives or nonrelatives, which maximizes the return on investment to their genes. To evaluate this idea, we derived a model for cooperative investment and tested its predictions using a social microbe that cooperatively builds a stalk to facilitate spore dispersal. We find that cooperative investment into stalk closely matches predictions, with strains strategically adjusting investment according to their relatedness to their group.

Introduction

Cooperation is widespread in nature (Hamilton, 1964b; West *et al.*, 2007a; Bourke, 2011), often being manifested as individuals investing in the production of public goods that benefit all members of a group (Hardin, 1968; Rankin *et al.*, 2007; Frank, 2010). However, these goods are vulnerable to exploitation by 'cheaters' (or 'free riders') that reap the benefits of cooperation without commensurate investment (Hamilton, 1963; Olson, 1965). Because such behavior has the potential to undermine the evolutionary stability of cooperation through public good production, successful cooperation is typically thought to require mechanisms of cheater avoidance or control (Clutton-Brock & Parker, 1995; Frank, 2003; Travisano & Velicer, 2004; West *et al.*, 2007a). This logic implies a simple evolutionary scenario where there is competition between alternative 'cooperator' and 'cheater' strategies. However, it is logical to assume that such discrete strategies would lose out to individuals that can strategically modify their contribution to public goods. This is because strategic investment could allow individuals to balance the costs and benefits of 'investing', whilst realizing potential opportunities to exploit the investments made by others (Ostrom, 1990; Doebeli *et al.*, 2004). Because these costs and benefits can vary across social settings, individuals face a strategic dilemma over how much to invest, with the realized success of a strategy depending not only on the level of cooperative investments made by the individual, but also that made by others in the group.

Kin selection theory provides an appealing framework for understanding how evolution shapes investment in cooperation. In this framework, the competing 'individuals' are different genetic variants (Williams, 1966; Cosmides & Tooby, 1981; Taylor *et al.*, 2007), with strategies evolving to maximize 'inclusive fitness' (Grafen, 2006a; West & Gardner, 2013). The inclusive fitness accounting considers the total impact of a behavior on the success of the causal genes in terms of the direct costs to the actor and indirect benefits to relatives (*i.e.*, others carrying that same genetic variant). For cooperation through production of public goods, where all benefits go to the entire group, relatedness to the group should be a critical

determinant of inclusive fitness because it governs the share of rewards that go to the individual, and hence determines the expected net return on investment. Consequently, we would logically expect that individuals should optimize their inclusive fitness by facultatively modulating their willingness to invest into public goods as a function of their relatedness to the members of the group (Hamilton, 1964a; Taylor & Frank, 1996; Pepper, 2000; Frank, 2010).

A number of theoretical studies have analyzed facultative cooperative strategies, where individuals modulate their behavior in response to social context (such as the behaviors shown by rivals) (Axelrod & Hamilton, 1981; Doebeli & Hauert, 2005). While most of these studies have focused on discrete alternative strategies ('cooperate' or 'cheat') (Axelrod & Hamilton, 1981; Doebeli *et al.*, 2004), there is also a growing literature that considers continuously variable strategic cooperative behavior in response to social contexts, including relatedness (Doebeli & Hauert, 2005; Frank, 2010). However, experimental tests of theoretical predictions often either rely on the simpler models that do not include such potential complexity (Sinervo & Lively, 1996; Dugatkin & Reeve, 1998; Turner & Chao, 1999; Bshary *et al.*, 2008; Gore *et al.*, 2009) or do not evaluate whether the observed facultative patterns are strategic (*i.e.*, match adaptive quantitative predictions from evolutionary models) (Buttery *et al.*, 2009; Manhes & Velicer, 2011; Parkinson *et al.*, 2011; Xavier *et al.*, 2011; Pollak *et al.*, 2016; Bruce *et al.*, 2017). For example, the opportunistic pathogen *Pseudomonas aeruginosa* facultatively produces iron-scavenging siderophores, which represent a cooperative public good (West & Buckling, 2003; Griffin *et al.*, 2004; Diggle *et al.*, 2007). Cells produce quorum-sensing molecules that that allow them to modulate their production of siderophores. There is evidence that investment into siderophore production is flexible (West & Buckling, 2003; Diggle *et al.*, 2007) and varies between broad-scale differences of 'high' versus 'low' relatedness (Griffin *et al.*, 2004). However, it is unclear as to whether the level of production can be varied quantitatively as a strategic response to fine-grained variation in relatedness.

To understand how selection shapes patterns of investment into public goods in response to variation in relatedness, we therefore first developed a dynamic game-theoretical framework that views competing genetic variants as players who can modulate their contributions to public goods based on their relatedness to their group. The resulting 'Collective Investment' game offers an intuitive economic logic for why and how organisms should modulate their contributions to public goods and provides a set of simple and unambiguous predictions that can be tested empirically. To directly test these predictions, we next examined the consequences of experimentally manipulating social group composition in the social amoeba *Dictyostelium discoideum* on patterns of individual and collective investment in cooperation. These studies revealed a remarkable agreement between patterns of individual and collective investment with fine-scale model predictions, where patterns of cooperation are explained by savvy investment strategies that maximize the fitness return on investment.

Results and Discussion

The Collective Investment game

When individuals engage in social interactions, their success typically depends on both their own behavior and the behavior of their social partner(s). Under these conditions, game theory provides a powerful framework for identifying how individuals should behave to maximize their expected social success across encounters (Maynard Smith & Price, 1973; Maynard Smith, 1974; Queller *et al.*, 1985; Taylor *et al.*, 2007). Game theoretical models predict that individuals will display the evolutionarily stable strategy (ESS), which cannot be invaded by any competing strategy (Maynard Smith & Price, 1973; Taylor *et al.*, 2007). In most economic and biological scenarios that involve cooperation, we might logically expect that individuals could do better by playing dynamic strategies in which they change their behavior quantitatively across different social contexts (Hamilton, 1964b; Ostrom, 1990).

While games with fixed alternative strategies (*e.g*., the Prisoner's Dilemma) have been widely used as the basis for analyses of strategic modulation of cooperative behavior (Doebeli & Hauert, 2005; Frank, 2010), they do not yield any quantitative predictions about continuously variable behavior. Instead, models that consider cooperation via public goods (Frank, 1995, 2010; Doebeli & Hauert, 2005; Dionisio & Gordo, 2006), typically based on the inclusive fitness framework (Taylor, 1992; Taylor & Frank, 1996; Frank, 1998), have proven more informative. We extend this work by developing a model based on an equivalent 'direct fitness' accounting, where different genetic variants are the players in a dynamic game, to provide an intuitive logic for the costs and benefits of investing in pubic goods. The game is described with two players, but logically extends to include more.

The Collective Investment game is based on a scenario in which the payoff to a player is determined by two opposing factors: the costs suffered from investing in the public good and the resulting benefits from public good availability (Figure 1A). From the perspective of the group, this antagonistic relationship between costs and benefits results in a scenario where group success is maximized at some intermediate level of collective investment whenever public good production is favored by natural selection (Figure 1B). Examples of this sort of scenario, where overall success is maximized at an intermediate level of investment, are well documented, ranging from economics to biology (Gordon, 1954; Parker & Maynard Smith, 1990; Foster, 2004; Doebeli & Hauert, 2005). However, the level of collective investment that maximizes group success (denoted Θ_G in the model) will typically differ from the level of personal investment that maximizes individual fitness (Frank, 2003). This is because individuals suffer the cost of investment, yet their payoffs are divided among the collective. Therefore, we expect individuals to implement selfish strategies that maximize their return on investment in terms of fitness, which must balance their personal costs with the return they receive through their influence on collective success (Olson, 1965; Ostrom, 1990). The relative magnitude of the costs and benefits together define the strength of selection (denoted

Γ in the model), which reflects the rate at which group success declines as investment deviates from the level that maximizes group success (*i.e.*, deviates from Θ_G).

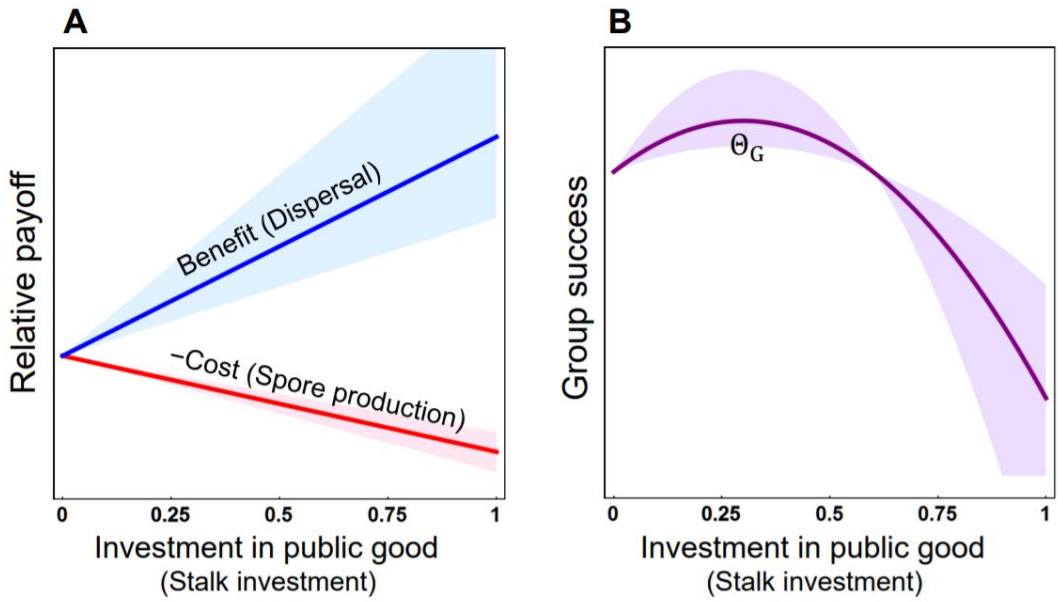


Figure 1. The costs and benefits of cooperation through production of public goods. **A)** The benefit (relative payoff) from public good production is an increasing function of the resources invested into the public good (blue line). Because investment is costly it results in a decreasing payoff through other components (red line). In the case of the *D. discoideum* system, the benefits of stalk investment come through spore dispersal and come at a cost in terms of reduced spore production. **B)** The costs and benefits of investment in the public good result in a quadratic relationship between total investment (I_G) and overall group success (ω_G). Groups have their highest success at some intermediate level of investment (Θ_G) that balances costs and benefits. In both A and B, investment in public good is given as the proportion of the total budget available, with zero being no investment and 1 corresponding to investment of all available budget into the public good. In the case of the *D. discoideum* system, this represents the proportion of cells that a strain invests into stalk production. For illustration, the optimal level of investment (Θ_G) resulting from the relative costs and benefits is 0.3. To capture different strengths of selection on investment (Γ, see equation 5), the bold lines were plotted for a strength of selection where Γ = 2, with the shaded region indicating the range from Γ = 1 to 4.

To implement our direct fitness accounting, we consider a player to represent some proportion of the group, which is equivalent to the frequency of that genetic variant within the group (and therefore can vary between 0 and 1) and represents their 'whole-group relatedness' (Taylor & Frank, 1996; Pepper, 2000) (in economic terms, this might be described as a player's 'stake' in the group). This measure of relatedness is relevant because, as the benefits of public goods are accessible to all group members, the whole-group is the beneficiary of investment made by an individual, and hence whole-group relatedness accounts for direct fitness return from investing in public goods. Despite differing from the more typical 'kinship' coefficient of inclusive fitness models, the two approaches produce exactly equivalent results (Taylor & Frank, 1996; Frank, 1998; Pepper, 2000). To identify the strategy that maximizes expected individual fitness, which represents the ESS for the game, we solved the Collective Investment game across the full range of relatedness over a broad array of relative costs and benefits of investment in public goods. These analyses revealed a general qualitative prediction for patterns of investment under the ESS: individuals should modulate their investment into public goods as a continuous function of their relatedness to the group. By evaluating the patterns predicted by the model across an enormous range of values for the optimal level of collective investment (*i.e.*, the value that maximizes group success, Θ_G) and the strength of selection on investment (Γ), it is clear that the qualitative results are robust across a wide array of conditions (Figure 2A and 2D; see also SI Appendix, Figure S1A, S1D, S1G and S1J). When there is a relatively large asymmetry in the degree to which players are related to the group, each player should behave differently. The player with higher relatedness to the group has the incentive to invest because their interests are more closely aligned with those of the group (and hence investing maximizes their fitness, see SI Appendix, Figure S2), while the player(s) that is less related to the group does best by withholding investment (or under-investing) and exploiting the investment of their partner (SI Appendix, Figure S2). Consequently, under these conditions, the player with the lower relatedness will have higher relative fitness than the player with higher relatedness because of this exploitative behavior (Figure 2B and 2E, see also SI Appendix, Figure S1B, S1E, S1H and S1K). In contrast, when

the players have similar levels of relatedness to the group, neither is expected to be willing to invest heavily, leading to a pattern of under-investment in the public good (Figure 2C and 2F; see also SI Appendix, Figure S1C, S1F, S1I and S1L).

Because organisms in nature presumably rely on some cue(s) to measure their level of relatedness to the group (which would represent a mechanism of kin discrimination), we also evaluated how the patterns would be affected if individuals make errors when measuring relatedness (with the patterns in Figure 2 and S1 illustrating the scenario of no measurement error). We included measurement error in the model by integrating over a Gaussian distribution centered on the true relatedness (allowing us to vary the degree of error by modulating the standard deviation of the error distribution, SI Appendix, Figure S3). We further assumed that measurement error depends on group complexity, and so is high at intermediate levels of relatedness (where group composition is the most complex), and low when one player has very high relatedness to the group. This extension of the model provides us with a robust and clear set of predictions for what to expect in nature (see Figure 3 for an example and SI Appendix, Figure S4 for illustrations across parameter space). Together, the Collective Investment game reveals that although the exact patterns will depend on the relative costs and benefits of public good production (which will determine the optimal level of investment and the relative strength of selection on investment patterns) and the degree of error in measurement of relatedness, the qualitative patterns of individual investment, relative fitness, and collective investment are consistent across parameter space (see also SI Appendix, Figure S2B for an illustration of absolute fitness).

A
B
C
D
E
F
Γ
¼
½
1
2
4
12
θG = 0.3
Γ = 2
θG
0.05
0.1
0.2
0.3
0.4
0.5
Investment in public good
(stalk investment)
Relative payoff
(relative fitness)
Collective investment
Relatedness
(Strain frequency in group)
θG
½θG
0
0.25
0.5
0.75
1
0.75
0.5
0.25

1

Figure 2. Examples of the predictions of the Collective Investment game (and specific application to the *D. discoideum* system). Predictions are plotted as a function of a focal player's relatedness to the group (*i.e.* a strain's frequency in the group). For parts A-C the optimal investment (θ_G) was fixed at 0.3 and the strength of selection (Γ) was varied, while for D-F the strength of selection was fixed at 2 and the optimum was varied. **A & D)** Predicted Investment ($I_{i|p_i}$) in the public good (stalk investment) as a function of relatedness (frequency). **B & E)** Predicted relative payoff (fitness) (ρ_i) as a function of relatedness (frequency). **C & F)** Predicted collective investment (I_G) for a pair of players as a function of the relatedness (frequency) of the focal player to the group. (See SI Appendix, Figure S1 for illustrations across other parameter values).

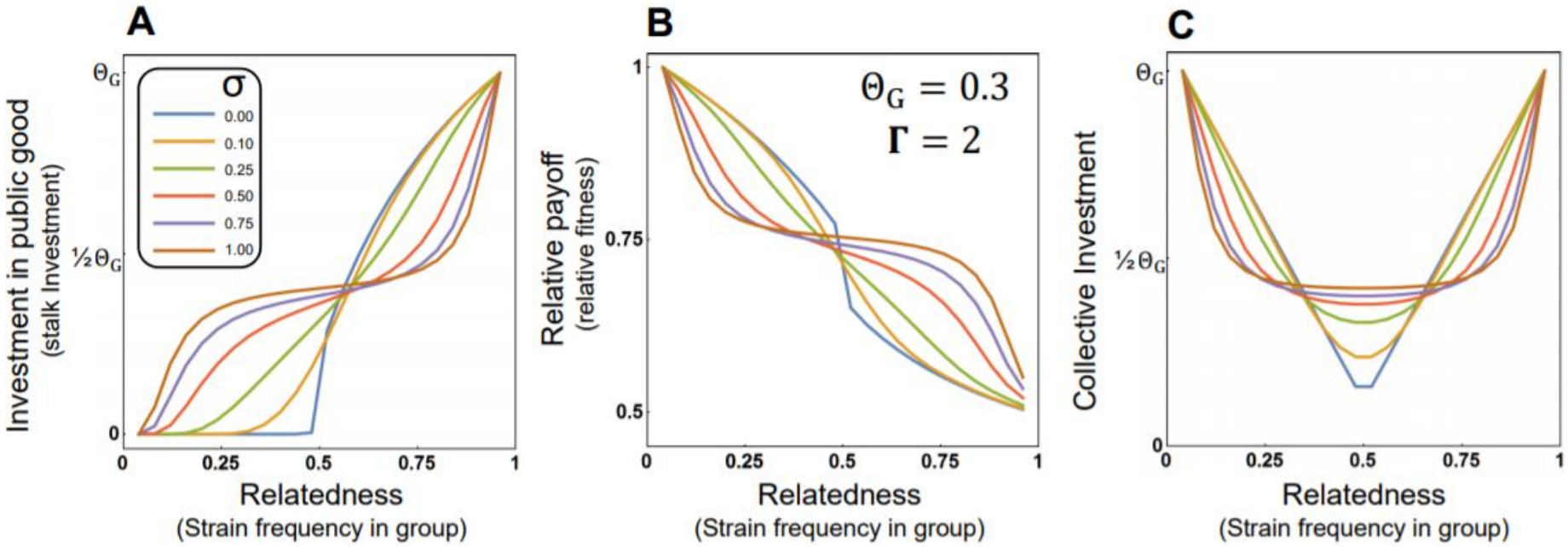

A
B
C
Investment in public good
(stalk Investment)
Relative payoff
(relative fitness)
Collective Investment
Relatedness
(Strain frequency in group)
σ
0.00
0.10
0.25
0.50
0.75
1.00
$\Theta_G = 0.3$
$\Gamma = 2$
Θ_G
½Θ_G
0
0.25
0.5
0.75
1

Figure 3. Illustration of the predictions of the Collective Investment game for the case where players make errors when measuring their relatedness. This corresponds to the scenario where players have imperfect information about their relatedness and are estimating their relatedness from some cues. The structure of the figure matches that of Figure 2. In all figures the optimal level of investment (θ_G) is 0.3 and the strength of selection (Γ) is 2. Lines within each figure correspond to different values of error (σ) in measurement of relatedness (frequency in the group) (see SI Appendix, Figure S4 for illustrations across other parameter values).

Individual and collective investment in D. discoideum

To test whether organisms are able to deploy the relatedness-dependent (and hence frequency-dependent) strategies predicted by the Collective Investment game, we measured patterns of investment into a public good in the social amoeba *D. discoideum*. Free-living *D. discoideum* amoebae initiate a social cycle in response to starvation (Strassmann *et al.*, 2000; Kessin, 2001). Thousands of amoebae aggregate to form a multicellular fruiting body with a supporting stalk composed of dead cells that holds aloft a sporehead. The stalk structure is thought to have evolved to aid spore dispersal, and it has been shown experimentally that an intact fruiting body does indeed increase dispersal (although dispersal is not eliminated by stalk removal) (Smith *et al.*, 2014). Stalk cell differentiation has typically been viewed as altruistic self-sacrifice for the benefit of the cells in the sporehead (Strassmann *et al.*, 2000, 2011; Foster *et al.*, 2004; Shaulsky & Kessin, 2007). However, this perspective ignores the implications of collective investment on the group's success: if a genotype only produced altruists then there would be no spores to reap the benefits of stalk investment and likewise, if a genotype only produced spores then they would be unable to reap group benefits of producing a stalk (Figure 1A). Consequently, there must be some intermediate level of stalk investment that is favored by natural selection that balances these costs and benefits (Figure 1B). Indeed, laboratory measurements reveal that typically 25-35% of cells are allocated to the stalk cell fate (Forman & Garrod, 1977; Chattwood *et al.*, 2013).

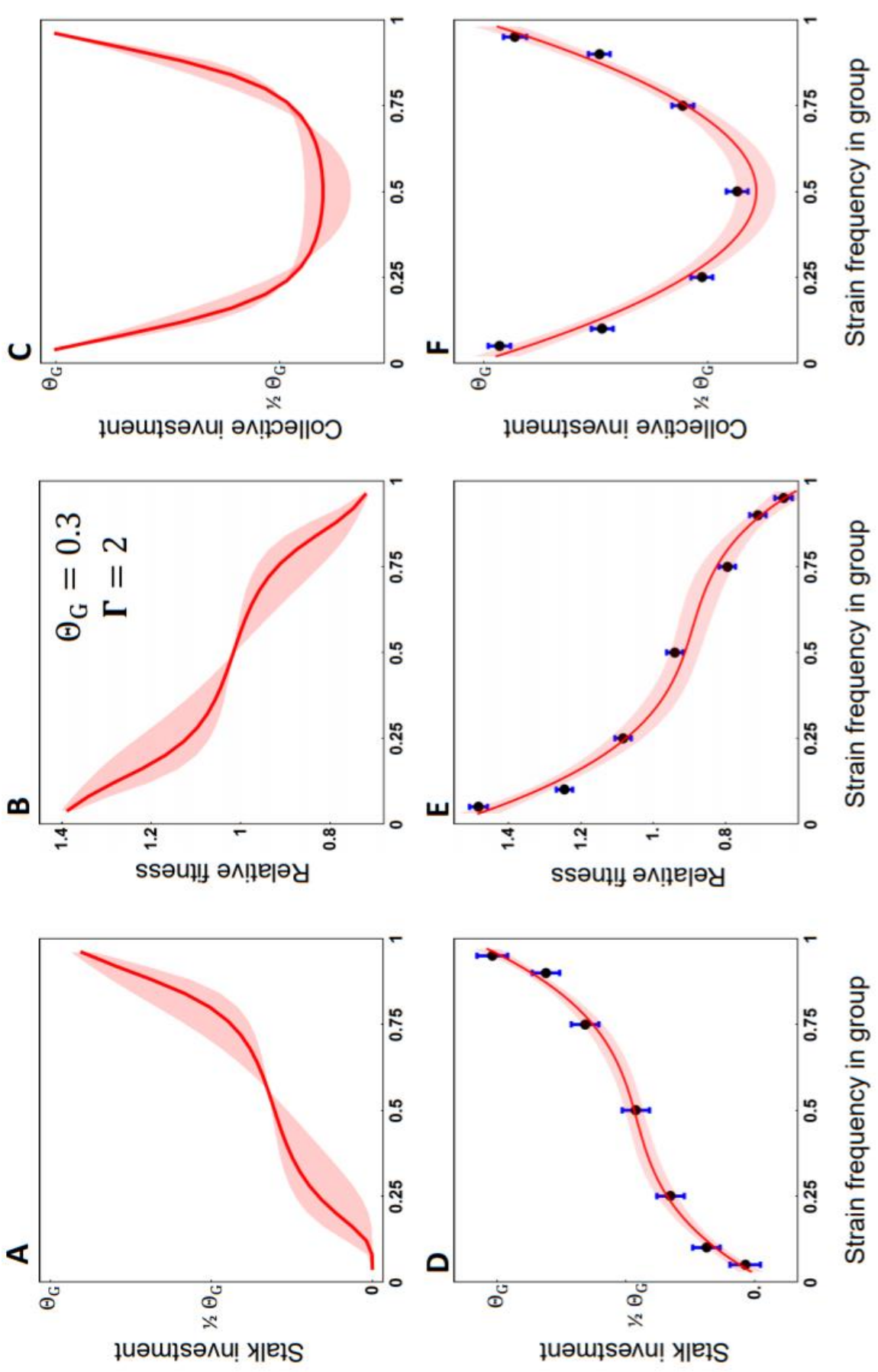
Multicellular aggregations can also be composed of multiple strains (*i.e.* can be chimeric), providing the opportunity for conflict over stalk investment (Strassmann *et al.*, 2000; Foster *et al.*, 2002). Conflict arises because the different strains within an aggregation each contribute to the costs for building the stalk, while all members of the aggregation benefit equally. Thus, stalk investment in *D. discoideum* fits the scenario modelled by the Collective Investment game. In our direct fitness accounting, different strains are the relevant fitness-maximizing strategists, with the proportion of cells sacrificed by the strain to build the stalk representing their investment into the public good, and their relative frequency within the

aggregation determining their relatedness to the group (Figure 1). Furthermore, *D. discoideum* provides an ideal model social system to experimentally test the predictions made by the Collective Investment game because group composition can be manipulated and corresponding patterns of investment can be measured quantitatively (Strassmann *et al.*, 2000). Specifically, the ESS of the Collective Investment game predicts that *D. discoideum* strains should show relatedness-dependent patterns of investment, meaning that their investment should change as a function of their frequency in a group. When a strain is at low frequency in the aggregation they would be predicted to invest little or nothing into the stalk (hence produce mostly spores), while a strain that is at a high frequency in an aggregation should invest at a level that is close to their clonal investment (Figure 2A and 2D). This pattern of investment results in a return on investment, and hence relative fitness, that is highest when a strain is at low frequency in an aggregation and hence has low relatedness (because it exploits its partner as a free rider) and is lowest when it is at high frequency and hence has high relatedness (because it pays the cost of being exploited). Consequently, the expected relative fitness of the lower-frequency player is always higher than that of the higher-frequency player (Figure 2B and 2E).

To test these predictions, we measured the behavior of co-occurring natural *D. discoideum* strains in clonal and chimeric development. We examined the fit to theoretical predictions using data from ten naturally co-occurring strains, which represent the spectrum of genetic diversity within a natural population (Gruenheit *et al.*, 2017), interacting in 34 different chimeric pairings. To vary levels of relatedness we combined pairs of strains across a range of frequencies (at least five different frequencies per replicate, for a total of 944 chimeric combinations). On average, strains show patterns of frequency-dependent investment in the stalk in pairwise mixes that match the qualitative predictions of the ESS in the Collective Investment game (compare Figure 4A with 4D, see also expected values in Figures 2A, 2D, and 3A). Strains invest little into the stalk when their relative frequency in a group is low and much more when their relative frequency is high ($\chi^2_{(3)} = 181.5$, $p < 10^{-38}$, see

also Figures S5A and S5B for high resolution illustrations of patterns from two pairings). Overall, the pattern very closely corresponds to the quantitative predictions of the model (Figure 4A and 4D). Strains approach zero investment when they are at a very low frequency in a group, whereas their investment is close to the optimal level of investment (assumed to be about 30% of their cells into stalk) when their frequency in a group approaches 100%. This pattern of investment leads to the pattern of frequency dependent relative fitness predicted by the Collective Investment game (Figure 2B, 2E, and 3B) in which strains have a high relative fitness when they are at a low frequency in a group and low relative fitness when they are at high frequency ($\chi^2_{(3)} = 348$, $p < 10^{-75}$; compare the illustration of expected values in Figure 4B with the experimental results in Figure 4E, see also Figures S5C and S5D). Importantly, these results imply that all strains will appear to behave as 'cheaters' when at low frequency in groups and as 'cooperators' when at high frequency.

The predictions of the Collective Investment game can also be viewed from the perspective of the aggregate behavior of the strains in terms of total collective investment. Experimental measurements of total collective investment as a function of the relative frequencies of strains shows the predicted pattern of relative investment in stalk across frequencies in a group (Figure 2C, 2F and 3C), where investment is lowest when strains are at the same frequency, and increases exponentially as the difference in their frequencies increases (*i.e.* as frequency of the focal strain approaches zero or one) ($\chi^2_{(2)} = 144.3$, $p < 10^{-32}$; compare the illustration of expected patterns in Figure 4C with empirical results in Figure 4F, see also S5E and S5F).


A
B
C
D
E
F
Stalk investment
Collective investment
Relative fitness
Strain frequency in group
$\Theta_G = 0.3$
$\Gamma = 2$
Θ_G
½ Θ_G
0
0.25
0.5
0.75
1
0.8
1.2
1.4

1

Figure 4. Patterns of stalk investment, relative fitness, and collective investment as a function of strain frequencies in chimeric aggregations. Parts A-C illustrate expected patterns (see Figure 3) under parameter values that resemble the empirical results (using the same equations [eqns. 8-10][1] to calculate model expectations as those used for empirical estimation), with the bold line corresponding to the case where Θ_G=0.3, Γ = 2, and $\sigma = 0.50$, with the shading spanning a range of error in measurement of frequency (relatedness) (σ = 0.25 to σ = 0.75). Parts D-F show empirical results from the set of 34 chimeric pairs (N=944 total chimeric mixes), with the points representing the means and the bars their standard errors, estimated from a mixed model (following the model structure in the Methods, but with frequency as a categorical factor). **D**) Individual stalk investment by a focal strain as a function of its frequency in a chimeric aggregation., **E**) Relative fitness for a focal strain as a function of its frequency in a chimeric aggregation, **F**) Collective investment by chimeras as a function of the frequency of a randomly assigned focal strain to the chimeric aggregation. In parts D and E the bold curve represents the best-fit estimate from the cubic regression model (here fitted to the estimated means). For part F, the curve represents the best-fit estimated from a quadratic regression model (fitted to the estimated means). For all three figures (parts D to F) the shaded region indicates a one standard error interval on either side of the best-fit line. For individual (parts A and D) and collective (parts C and F) investment values were re-scaled by subtracting $1 - \Theta_G$ from the raw measures, under the assumption that $\Theta_G = 0.3$ (therefore, the value labelled as Θ_G corresponds to a value of 0.3 in the figure).

[1] These equations were given incorrectly in the published manuscript as "*[eqns. 11-13]*". I have corrected the text to the correct equation numbers.

The Prisoner's Dilemma and how to avoid it

Although the ESS is characterized by continuously variable relatedness-dependent (or frequency-dependent) behavior (Figure 2), to achieve a more intuitive understanding we can link the payoff structure at any particular group composition to canonical games. To do so, at a given group composition we can compare the relative payoffs to a player that defects by making no contribution and the relative payoffs to a player that cooperates by making a contribution (see Methods). We consider a scenario to be akin to the Prisoner's Dilemma when defection is the best strategy for both players, regardless of the opponent's strategy. In the Snowdrift game, we expect players to adopt opposite roles, with one cooperating and the other defecting. Therefore, we consider two different scenarios to be akin to the Snowdrift game. The first scenario follows the structure of the classic symmetrical game, where players are better off defecting against a cooperator and cooperating against a defector. The second scenario occurs when there is an asymmetry between players that dictates their roles in the Snowdrift game, with one player doing best by cooperating while the other does best by defecting.

The exact nature of payoffs depend on the model parameters, but in general, when the players' differ widely in their relatedness to the group, we find that the pattern of joint payoffs are akin to the Snowdrift game and when they have similar levels of relatedness to the group it is akin to the Prisoner's Dilemma (Doebeli *et al.*, 2004; Doebeli & Hauert, 2005) (Figure 5). Under the Snowdrift game, one player adopts the role as the cooperator and the other as the defector, which results in relatively high fitness for the group. By adopting different roles, the defector receives a higher payoff than the cooperator, but the cooperator is willing to adopt that role because it is better off cooperating than defecting when its opponent defects (Doebeli & Hauert, 2005). In the context of the Collective Investment game, it is the asymmetry in relatedness to the group that drives the players to adopt the two roles (Figure 5), with the player that is more related to the group acting as the cooperator while its opponent is able to defect (Figure 2A, 2D and 3A), leading to a higher relative payoff to the defector (Figure 2B,

2E and 3B). In contrast, under the Prisoner's Dilemma conditions (Figure 5), both players do best by defecting, which leads to low collective investment (Figure 2C, 2F and 3C). These game scenarios help explain the pattern of collective investment in stalk that we observe in the *D. discoideum* system (Figure 4F): under the Snowdrift game conditions we see collective investment approach the level seen in clonal development (which presumably evolved to maximize group fitness), whereas under the Prisoner's Dilemma conditions we see underinvestment.

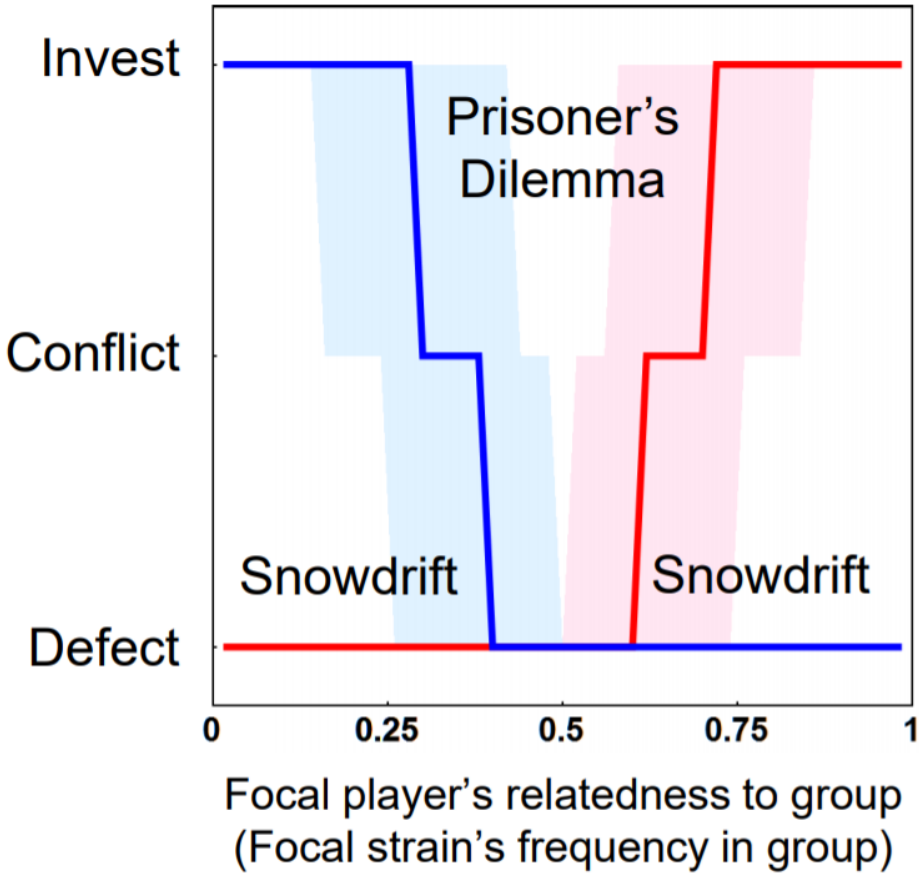


Figure 5. Payoff structure of the Collective Investment game and relationship to classic games. Payoffs are characterized in terms of whether defection or investment is favored, or whether the best strategy depends on the investment by the opponent (labeled as 'Conflict'). The best strategy for the focal player is shown in red and that of their opponent in blue. When both players do best by defecting the overall payoff structure is akin to the Prisoner's Dilemma, and we see low levels of total investment (see Figure 2). When one player does best by investing while its opponent does best by defecting the overall payoff structure is akin to an asymmetric Snowdrift game, where the difference in relatedness determines which player takes the role as the cooperator (with the player with higher relatedness making the investment in cooperation). Bridging these two regions is a zone of conflict. The bold lines correspond to a level of investment of $\frac{1}{2}\Theta_G$, with the shaded region spanning the range from $\frac{1}{4}\Theta_G$ to $\frac{3}{4}\Theta_G$. The shaded region illustrates that the zones corresponding to the different games will depend on how much an individual invests when cooperating.

We expect the predicted collective underinvestment under the Prisoner's Dilemma conditions to be detrimental compared to the higher investment under Snowdrift conditions. We tested this by measuring the proportion of fruiting bodies that collapsed due to inadequate investment in the stalk. Fruiting bodies made by chimeric mixtures (using all pair-wise 50:50 mixes of ten natural strains) were found to have spontaneously collapsed more often than those made by clonal groups (12% versus 1.1%, $F_{(1,52.3)} = 10.4$, $p = 0.002$). Furthermore, we expect the stability of fruiting bodies to reflect the overall level of collective investment in stalk, which should be manifested as an inverse relationship between the level of collective investment (Figure 4F) and the probability of fruiting body collapse. We tested this prediction using data from four pairs of strains measured at seven frequencies and find the expected negative correlation between collective investment for a given pair and probability of their fruiting bodies collapsing ($r = -0.94$, $p = 0.0009$). This relationship between investment and fruiting body stability underlies a strongly frequency (and hence relatedness) dependent risk of fruiting body collapse, with risk of collapse peaking when is there is no asymmetry in the frequency of the strains (*i.e.* both strains at a frequency of 0.5) and declining exponentially as the difference in frequencies increases (*i.e.* on either size of a frequency of 0.5) ($\chi^2_{(4)} = 403$, $p < 10^{-86}$; Figure 6A, Figures S6 and S7). If we use fruiting body stability (which is simply 1 minus the probability of fruiting body collapse) as a proxy for the dispersal success of a group ($\phi_{dispersal(G)}$) and the estimates for individual stalk investment (see Figure 4D) to estimate fitness through spores (as simply 1 minus the proportion of cells invested in stalk; see eqn. 1), we can generate an approximate pattern of individual fitness (see eqn. 3). Despite the fact that our lab-based measure of fruiting body stability provides only a rough approximation for group fitness through dispersal, we find that the pattern of individual fitness closely matches the pattern expected under the Collective Investment game (SI Appendix, Figure S2C and S2D). The resulting fitness pattern illustrates that individuals will have the lowest possible fitness when at intermediate frequencies and, while individuals always do best at very low frequency in a group, individual fitness increases towards both frequency extremes.

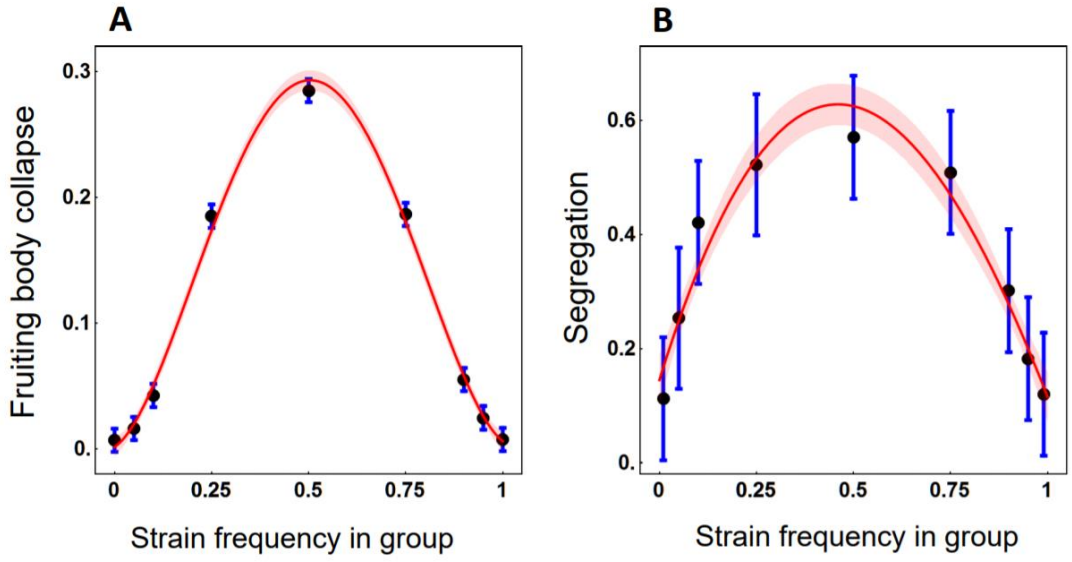


Figure 6. Empirical measures of fruiting body stability and segregation behavior. **A)** The proportion of fruiting bodies that spontaneously collapsed as a function of the frequency of the focal strain in each mix (estimated from six chimeric pairings; N=324). **B)** The relative degree of segregation as a function of the frequency of the designated focal strain. Measurements are from three different chimeric pairings across the nine frequencies (N=692 total sporeheads, with an average of 25.6 sporeheads measured for each pair at each frequency). In both figures, the points represent the means and the bars their standard errors, estimated from a mixed model (following the model structure in the Methods, but with frequency as a categorical factor). For part A, the curve gives the best fit cubic relationship while for part B the curve gives the best fit quadratic relationship (with the shaded region indicating a one standard error range on either side of the curve).

The finding that individuals suffer a much larger cost from conflict when trapped in the Prisoner's Dilemma-like conditions at intermediate levels of relatedness (Figure 5 and 6A) raises the question of why strains would engage in cooperative fruiting body formation under these conditions. Indeed, widespread (imperfect) strain segregation is a known mechanism in *D. discoideum* for avoiding chimerism when strains are mixed at equal frequencies and developed on a natural soil substrate (Benabentos *et al.*, 2009; Gilbert *et al.*, 2012), with two rapidly evolving genes being thought to be principally responsible (Benabentos *et al.*, 2009; Gruenheit *et al.*, 2017). Although the mechanism by which these genes regulate segregation remains to be fully elucidated, there is evidence to suggest that a critical mass of self-self-interactions are required for the coordinated directional motility that is necessary to form independent cooperating groups (Ho & Shaulsky, 2015). We might, therefore, expect strains to only show segregation when faced with Prisoner's Dilemma-like conditions (*i.e.* low asymmetry in relatedness), while remaining in aggregations when in Snowdrift-like conditions of high asymmetry in levels of relatedness. Indeed, as predicted, we find that segregation is highest when there is little asymmetry in frequencies (relatedness) and declines exponentially as the difference in frequencies increases ($\chi^2_{(2)} = 19$, $p < 10^{-4}$; Figure 6B). The frequency-dependent nature of segregation suggests that it may not have evolved as a mechanism of 'cheater avoidance', as has previously been suggested (Benabentos *et al.*, 2009; Gilbert *et al.*, 2012; Gruenheit *et al.*, 2017), but rather, as a mechanism for reshaping group composition to generate asymmetry in relative frequencies (resulting in a scenario where there will typically be a strain with high relatedness to the group), thereby avoiding the pernicious Prisoner's Dilemma-like conditions and entering into the more favorable Snowdrift-like conditions.

The logic of collective investment

The Collective Investment game and the supporting empirical data from the *D. discoideum* system have broad implications for our understanding of cooperative behavior. From the perspective of kin selection theory, an individual's relatedness to the group governs

whether the personal cost of contributing to public goods are outweighed by the benefit. Consequently, if individuals can measure their relatedness to group-mates, we would expect to see them invest in a way that maximizes inclusive fitness in terms of the balance between the benefit to kin in relation to the costs to self (following Hamilton's rule in the context of the ESS, which means that the optimal strategy depends on the behavior of opponents). Applying this logic to the *D. discoideum* system, an individual cell should modulate its 'willingness' to differentiate into a stalk cell based on its measurement of relatedness to other members of its aggregation, with the actual level of investment being determined by the benefits of producing a stalk relative to the cost of diminished spore production. From an economics perspective, we can view players as investors in some collective venture who are out to maximize return on investment, with relatedness representing their level of 'stake' or 'ownership' in the venture. When a player has a low relatedness to the group, their personal investment can have little effect on the overall performance of the venture (regardless of how much they invest), so they are better off withholding their investment. In contrast, when a player has high relatedness to the group their investment can have a large impact on the performance of the venture. Therefore, they should be willing to invest more heavily. In the context of the *D. discoideum* system, this perspective logically implies that a strain at a low frequency in an aggregation cannot impact the performance of the fruiting body regardless of how much it invests into stalk, and hence that strain should withhold their investment. Finally, we can view the scenario from the perspective of a dynamic game, with individuals as players out to maximize their payoff. From this perspective, a player contributes to the public good because they directly benefit from their own contribution and the optimal strategy is determined by the benefit they receive in relation to the cost paid (see Figure 1A). Players with low representation in the group do not contribute much to the public good because their contribution is diluted by the group, so they receive back only a small fraction of what they invest. In contrast, a player with high representation in the group should invest more because they receive back most of the benefit, and consequently they are mostly helping themselves through production of the public good. In the context of the *D. discoideum* system, this

perspective implies that a strain with a high frequency in an aggregation should contribute heavily to stalk production because most of the benefit goes to their own spores, and lower investment would only hurt them. The result is that a strain with a lower frequency, who would see little return on their contribution, can be exploitative since it is in the best interests of a common strain to build a stalk to their own benefit.

Although these different perspectives suggest different logical explanations for why and how individuals should invest in public goods, they are ultimately interchangeable since all are based on the same underlying framework. All suggest that organisms should adopt dynamic strategies in which they modulate their contribution to cooperation through public goods in relation to their relatedness to the group. Furthermore, it suggests that approaches where organisms are simply classified as 'cooperators' and 'cheaters' (Maynard Smith, 1976; Doebeli *et al.*, 2004; Travisano & Velicer, 2004) will often fail to capture the true nature of cooperative behavior in many systems. Indeed, the same individual or genotype could be expected to be cooperative or exploitative depending on their relatedness to the group. This scenario is clearly realized in the *D. discoideum* system. Although strains have typically been viewed as cooperators and cheaters (Strassmann *et al.*, 2000; Gilbert *et al.*, 2007; Strassmann & Queller, 2012; Santorelli *et al.*, 2013), the striking fit of the observed investment behavior by natural strains to the predictions of the Collective Investment game (Figure 4) provides strong evidence they cooperative through the implementation of a dynamic frequency-dependent strategy. As a result, all strains can appear as cheaters when they are at a relatively low frequency in a group and as cooperators when they are at a relatively high frequency. Our finding that even simple organisms like a social amoeba can implement the sorts of savvy relatedness-dependent investment suggests that these dynamic adaptive strategies may be common in nature.

Materials and Methods

The Collective Investment game

The Collective Investment game is a two-player game in which each individual makes an investment into a public good and receives a payoff as a function of their own investment and the collective investment of the pair. The structure of the game is related to economic games of public goods (Olson, 1965; Frank, 2010), but differs in that the return on investment is a function of a player's relatedness to the group. The game is described with reference to the *Dictyostelium discoideum* system but the basic structure is easily adapted for other systems. The players are different genotypes (strains), but in principle can represent any evolutionarily-relevant fitness-maximizing agent. Within an aggregation (which represents the group or collective) strains may have varying relative frequencies or proportions (p_i). The frequency of a strain in a group is equivalent to whole-group relatedness since it represents the average relatedness of a randomly selected cell to the entire group (self-included) (Taylor & Frank, 1996; Pepper, 2000). We present the model results and insights with regard to relatedness in keeping with theory but discuss the results in the context of frequencies of strains within a group to provide a clear link to the experimental methods.

Strains invest a proportion of their cells into stalk ($I_{i|p_i}$) and the rest ($1 - I_{i|p_i}$) into spores (with the level of investment potentially depending on their proportion, p_i). Therefore, their level of investment represents the proportion of their entire 'budget' of cells that are allocated towards stalk production (hence $0 \leq I_{i|p_i} \leq 1$). Investment into stalk is costly because it reduces the total number of spores a strain can produce and hence the 'payoff' (component of fitness) to a strain through spores declines (at a rate of γ_s) as a function of their investment in stalk (see Figure 1A):

$$\phi_{spores(i)} = 1 - \gamma_s I_{i|p_i} \quad (1)$$

The payoff is scaled to a value of 1 when no cells are invested into stalk.

Strains presumably invest in building a stalk to facilitate dispersal of spores (Strassmann *et al.*, 2000; Foster *et al.*, 2002; Smith *et al.*, 2014). While the cost of investing into the stalk is paid by the individual strain from their total budget of cells, the benefit (payoff) gained from dispersal depends on the architecture of the fruiting body, and hence on collective investment into the stalk (which is simply the weighted average of the stalk allocation of the two players, $I_G = \sum I_{i|p_i} p_i$). We model the performance of the fruiting body for spore dispersal as an increasing function of collective investment:

$$\phi_{dispersal(G)} = 1 + \gamma_d I_G \quad (2)$$

where γ_d gives the rate at which the payoff through dispersal increases as a function of investment into stalk (Figure 1A). As with the payoff through spores (eqn. 1), the payoff through dispersal is scaled to a value of 1 when no investment is made. For both payoff functions (eqns. 1 and 2) the qualitative results do not depend on this scaling, so the baseline value of 1 is used in both cases for simplicity. Similar cost/benefit relationships underlie a wide array of models that consider tradeoffs, such as models for the evolution of life-histories (*e.g.*, models of clutch size and parental investment). For example, models for the evolution of parental investment assume increasing investment per offspring is costly because it reduces fecundity, but beneficial because it increases offspring survival. Although, like many of these models, we assume a linear relationship between investment and costs/benefits, the qualitative results are robust across an array of relationships (so long as costs and benefits both increase with investment).

The overall success of a strain is determined by its payoff through spores weighted by the overall performance of the fruiting body. This is consistent with evolutionary theory, such as models that consider trade-offs between components of fitness or episodes of selection, and

is necessary to properly account for the influence of multiple factors affecting fitness. For example, to calculate total parental fitness in models for the evolution of parental investment, it is necessary to multiply an individual's fecundity (number of offspring produced) by the expected survival of the progeny they produce (since the product represents the number of surviving offspring). In terms of the *D. discoideum* system, the expected success of each spore depends on its expected dispersal, and hence fitness of a strain is the product of spore number and spore dispersal:

$$\omega_i = \phi_{spores(i)}\phi_{dispersal(G)} \quad (3)$$

The overall success of a group is simply the average fitness of its members (eqn. 3), $\omega_G = \sum \omega_i p_i$ which is equivalent to the expected payoff for the group through spores weighted by the payoff through dispersal, $\omega_G = \phi_{spores(G)}\phi_{dispersal(G)}$ (where the group payoff through spores is the weighted average of the spore production by group members, $\phi_{spores(G)} = \sum \phi_{spores(i)} p_i$). The trade-off between spore production and spore dispersal reflected in the payoffs (eqns. 1 and 2, see Figure 1A) results in a quadratic relationship between collective investment and group success (Figure 1B). From this relationship, we can derive the level of collective investment (I_G) that maximizes group success ($I_G = \Theta_G$), which represents the most efficient (welfare optimal) allocation of cells to stalk and spores that is possible given the costs and benefits of stalk investment:

$$\Theta_G = \begin{cases} \frac{1}{2}\left(\frac{1}{\gamma_s} - \frac{1}{\gamma_d}\right), & \text{if } \left(\frac{1}{\gamma_s} - \frac{1}{\gamma_d}\right) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where the condition insures that investment is non-negative. Therefore, the optimal level of investment into stalk (in terms of group success) is determined by the relative importance of payoffs through spores versus through dispersal. Consequently, under any conditions where the benefits of dispersal outweigh the cost to spore production, the collective will have highest overall success at some intermediate level of investment into stalk. Because aggregations of *D. discoideum* invest into stalk while also producing spores, the pattern of payoffs in nature must result in such an intermediate optimum. The strength of selection on fruiting body architecture (Γ) is given by the rate at which group success declines as the level of investment deviates from the group optimum:

$$\Gamma = -\gamma_s \gamma_d \tag{5}$$

The value of Γ represents the curvature of the relationship between collective investment and group success (*i.e.*, it is the quadratic coefficient for the parabolic relationship between collective investment and group success; see Figure 1B).

While equation (4) represents the optimal investment into stalk for a group, individual players (strains) within a group should invest in a way that maximizes their expected individual fitness (eqn. 3). The optimal level of investment for a given player (a strain) is a function of their relatedness to (*i.e.*, frequency in) their group:

$$\Theta_i = \begin{cases} \frac{1}{2}\left(\frac{1}{\gamma_s} - \frac{1}{\gamma_d p_i} - \frac{I_{j|p_j} p_j}{p_i}\right), & \text{if } \left(\frac{1}{\gamma_s} - \frac{1}{\gamma_d p_i} - \frac{I_{j|p_j} p_j}{p_i}\right) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Logically, the optimal level of individual investment corresponds to the value that maximizes group success (eqn. 4) when a strain is clonal ($p_i = 1$). At all other frequencies, the optimal level of investment will be lower than the value that maximizes group success (since $0 \geq I_{j|p_j}$

≤ 1 and $0 > p_i < 1$). The level of investment given by equation (6) represent the ESS for a strain, but because the optimal level of investment by each strain depends on the level of investment by other strains, the actual level of investment will depend on the joint resolution of that interdependence. As a result of this interdependence, the constraints on the range of investment values ($0 \leq I_{i|p_i} \leq 1$), and the constraints on the range of frequencies ($0 \leq p_i \leq 1$), we use numerical solutions from equation (6) to illustrate the patterns of the ESS under different conditions (see below).

To understand the properties of the ESS consider the case where other strains make no investment, such that the ESS is simply $\frac{1}{2}(1/\gamma_s - 1/\gamma_d p_i)$ (or zero when the term is negative). This level of investment represents the most economically 'efficient' strategy for a strain. Under these conditions, when the optimal strategy is to make a non-zero stalk investment, the two terms in parentheses must be greater than zero, with the first term ($1/\gamma_s$) representing the reciprocal of the cost of investing and the second term ($1/\gamma_d p_i$) the reciprocal of the benefit of investing. Thus, at the optimal payoff $p_i\gamma_d > \gamma_s$, which is a form of Hamilton's rule (Hamilton, 1964a; Charnov, 1977), the kin selection benefits ($p_i\gamma_d$) must outweigh the costs (γ_s). The third term in parentheses ($I_{j|p_j}p_j/p_i$) reflects the dispersal benefit to the focal strain arising from investment into stalk made by other strains, with the numerator ($I_{j|p_j}p_j$) representing the total investment made by others. The ESS deviates from the most efficient strategy because any investment made by other strains increases the value of the focal strain's spores, and hence increases the cost of making their own investment. This term can be viewed from an economic perspective as an 'opportunity cost', where a strain has the opportunity to gain from the dispersal benefit provided by the investment made by others and loses that opportunity when those spores are sacrificed to invest into stalk. The kin selection consequences of this opportunity cost can be seen by examining the conditions where the ESS level of investment is non-zero, which correspond to $p_i\gamma_d > \gamma_s\left(1 + \gamma_d I_{j|p_j}p_j\right)$. Consequently, if we view these conditions as a form of Hamilton's rule, we can see that the

dispersal benefit to kin from investing has to overcome both the direct cost from making an investment and the additional cost arising from the missed opportunity to exploit investments made by others.

We can also view the cost of investment into stalk in terms of its effect on the representation of a strain in the sporehead of their group (p_i'), which defines their within-group fitness. Their representation is determined by their investment in stalk relative to the overall investment made by the group: $p_i' = p_i\left(1 - I_{i|p_i}/1 - I_G\right)$. The within-group fitness can be calculated as a strain's representation in the sporehead relative to its frequency in the group: $\widehat{\omega}_i = p_i'/p_i$, making the relative (within-group) fitness of a strain ($\rho_i = \widehat{\omega}_i/\widehat{\omega}_j$):

$$\rho_i = \frac{1 - I_{i|p_i}}{1 - I_{j|p_j}} \tag{7}$$

Therefore, relative fitness within a group is a direct function of the relative investment made by strains. The pattern of relative fitness within a group is similar to the pattern of relative absolute fitness (ω_i/ω_j), which is simply $\left([1 - \gamma_s I_{i|p_i}]/[1 - \gamma_s I_{j|p_j}]\right)$.

The nature of the game

To understand the properties of the ESS we can characterize the payoffs to players in relation to the payoff structures of the Prisoner's Dilemma and Snowdrift games (Doebeli & Hauert, 2005). This analysis allows us to relate the game's properties to the intuitive framework of existing well-understood models. However, to achieve this goal we need to first address the fact that the Investment Game differs from the canonical games in three key aspects. Firstly, the Investment Game differs in that expected payoffs vary as a function of relatedness, so there is no single payoff matrix, but rather, a relatedness-dependent payoff function. Therefore, we need to evaluate the properties of the game across levels of

relatedness, which allows us to understand how the properties of the game change as a player's relatedness to the group changes. Secondly, when the opposing players differ in their relatedness to the group, they will also differ in their expected payoffs. Therefore, we need to consider a separate payoff matrix for each player at each level of relatedness. Finally, because investment into public goods can vary quantitatively, the game does not have discrete strategies that correspond to fixed alternative strategies like 'cooperate' or 'defect'. There are several logical alternative ways to consider cooperation versus defection and the type of game that a scenario corresponds to necessarily depends on the level of investment being made by a 'cooperator'. The higher the investment made by a cooperator the higher the rewards for defection, which changes the optimal response (see eqn. 6). Therefore, we use a simple framework where we consider defection as the case where individuals make no contribution to the public goods and cooperation as the case where individuals make some non-zero contribution (the size of which we vary in our analysis of the game).

The game scenario depends on payoffs to a player in terms of their expected fitness (ω_i, eqn. 3) under four scenarios (stating the focal player's strategy first): cooperate against a cooperator (C_iC_j), cooperate against a defector (C_iD_j), defect against a cooperator (D_iC_j), or defect against a defector (D_iD_j). Because we are primarily interested in how payoffs lead to 'motivation' for a player to invest or defect, we consider 'weak' forms of the games rather than the overall structure of the payoff matrices. That is, we consider whether a player's fitness is increased or decreased by making a contribution to public goods when their opponent either cooperates (makes a contribution) or defects (withholds their contribution). Payoffs are classified as being Prisoner's Dilemma-like when a player is better off defecting regardless of the strategy of their opponent ($D_iC_j > C_iC_j$ and $D_iD_j > C_iD_j$) and Snowdrift like when they are better off defecting against a cooperator and cooperating against a defector ($D_iC_j > C_iC_j$ and $C_iD_j > D_iD_j$). If a player is better off cooperating regardless of the strategy of their opponent ($C_iC_j > D_iC_j$ and $C_iD_j > D_iD_j$) we consider their strategy as selfish investment, meaning they

are favored to cooperate because it is in their own selfish interests regardless of what their opponent does.

Both players can 'agree' on the game being played or, because of the asymmetry in payoffs, they can disagree. When both agree that the game is Prisoner's Dilemma or Snowdrift we classify the scenario as the agreed game. Disagreement over the game being played generally arises when one player views the scenario as favoring selfish investment, while the other sees the scenario as a Prisoner's Dilemma. This scenario is analogous to the ESS for an asymmetrical Snowdrift game, with one player getting a payoff for cooperating with a defector and the other getting the payoff for defecting against a cooperator. In this case, the asymmetry in relatedness determines which player will take the role as cooperator and which as defector (with the higher relatedness player being the cooperator). Hence, we describe this scenario as being like an asymmetrical Snowdrift game.

Imperfect information

The derivation of the Investment Game implicitly assumes that players (strains) have perfect information about their relatedness to the group and can therefore adjust their investment accordingly. In the context of *D. discoideum*, 'information' is the output of any mechanism that provide feedback to cells that reflects their frequency in a group, and hence can potentially arise from many molecular mechanisms. Of course, if the players have no information about their relatedness we would not expect to see any relatedness-dependent changes in stalk investment, so any frequency dependent change in behavior must correspond to some information (regardless of whether it is actively or passively acquired). Presumably any molecular mechanism or responses to information should have some degree of noise, resulting in random error in the measurement of relatedness. In the *D. discoideum* system, random noise could simply represent the variation from cell to cell in their measurement of their frequency, so the entire group of cells from a strain measures their frequency with some noise. The mean of their measurement could be accurate, but the individual cells would

respond as if they were at a different frequency, making the response deviate from the perfect information case.

We modelled error using a Gaussian probability density function (PDF), where the mean of the PDF represents the true frequency (relatedness) of the strain and the standard deviation the level of noisiness (see SI Appendix, Figure S2). We assume that measurement error depends on the complexity of group composition, so the magnitude of the error (*i.e.*, the standard deviation of the PDF) was weighted by $4p_1p_2$ (which has a maximum value of 1 when $p_1 = p_2$ and declines to zero as either strain nears a frequency of 1). Logically, this implies that strains are much more able to measure their frequency (relatedness) when they are at extreme frequencies than when they are at intermediate frequencies in a group. For example, a strain would be better able to distinguish between a true frequency of 0.01 and 0.21 than it would be able to distinguish between 0.4 and 0.6. Analyses were integrated over all possible frequencies (from zero to one), with the probability that a strain behaves as if it has a particular frequency being given by the PDF weighted by the group complexity term. Because each player assesses their own frequency, analyses at a given frequency require integration over all possible pairwise frequencies.

Model predictions

To generate predictions for collective investment in *D. discoideum*, we varied the relative cost to spore production (γ_s) and benefit from dispersal (γ_d) from stalk investment to alter the strength of selection on fruiting body architecture (eqn. 5). For most illustrations in the main text we restricted the parameters to values that result in an optimal level of clonal investment of 30% of cells to the stalk, which corresponds to the approximate pattern observed in naturally derived strains (Forman & Garrod, 1977; Chattwood *et al.*, 2013). However, in Figures 2D, 2E and 2F, we hold the strength of selection constant (at $\Gamma = 2$) and vary the optimal level of clonal investment to illustrate the impact of different optima. We illustrate a much wider range of parameter space in SI Appendix, Figure S1, varying both the strength of

selection and the clonal investment optimum systematically across panels. Within the range of values that keep fitness non-negative, the strength of selection on allocation of cells and the clonal investment optimum (which are both determined by the values of γ_s and γ_d, see eqns. 4 and 5) do not change the qualitative predictions of the model.

At each frequency (relatedness) we solved the ESS level of investment (eqn. 6) for the two players. Exact solutions were generated using the Solve function in Mathematica 10.0 (Wolfram Research, Inc.). Given the ESS level of investment, we calculated absolute and relative (within-group) fitness of each player and the level of collective investment. We also analyzed the game scenarios under each scenario to link these patterns to the logic of the Prisoner's Dilemma and Snowdrift games. To link the model results to the experimental data we also calculated individual and collective investment following the methods used in the experimental work (where all measures are based on spore counts and representation in chimeric sporeheads, see below).

Measurement of spore allocation

We followed well-established *D. discoideum* protocols (Forman & Garrod, 1977; Kessin, 2001; Buttery *et al.*, 2009), which are therefore only briefly outlined here. We used a set of ten naturally co-occurring strains of *D. discoideum* from Little Butt's Gap, North Carolina (NC28.1, NC34.2, NC52.3, NC60.1, NC63.2, NC69.1, NC71.1, NC80.1, NC99.1 and NC105.1) that have previously been used in several studies of social interactions (Buttery *et al.*, 2009; Wolf *et al.*, 2015; Gruenheit *et al.*, 2017). All strains were grown on SM plates with *Klebsiella aerogenes* as a food source. Before aggregation, cells were harvested and washed of bacteria by repeated centrifugation in KK2 (16.1mM KH_2PO_4, 3.7mM K_2HPO_4). To construct experimental chimeras, we reciprocally mixed cells from a strain that was fluorescently labelled with 10 mM CellTracker™ Green CMFDA dye with an unlabelled partner treated with DMSO to control for any effect of labelling. Clonal sets of labelled and unlabelled cells were also created to provide a measure of any counting bias. Cell mixes were

plated for development on 1.5% KK2 purified agar plates (surface area ~21.3 cm^2), at a density of $4.7x10^5$ cell/cm^2. Relative proportional representation of the focal strain in the sporehead was primarily determined by counting the percentage of fluorescent spores using flow cytometry. However, for some sets of replicates from two pairs (NC28.1+NC63.2 and NC34.2+NC105.1) measurements were done by microscopy (with spores washed into 5ml spore buffer and imaged using a fluorescence imaging system). Despite the fact that two different methodologies were used to measure relative spore number, the patterns of relative representation in the sporehead were indistinguishable. Because of technical limitations associated with the labelling process, an average of 0.3% (s.d. = 0. 09%) of unlabelled spores are counted as being labelled and an average of 1.4% (s.d. = 0. 9%) of labelled spores are counted as being unlabelled (based on data from clonal populations of labelled and unlabelled). Therefore, to correct for any potential counting bias, the raw proportion of labelled (p_i^*) cells of strain *i* in a chimeric mix with an unlabeled strain *j* was corrected using the proportion of labelled cells measured from clonal sets of labelled ($p_{i(C)}^*$) and unlabelled cells ($p_{j(C)}$) (created using the same pools of cells as in the chimeric mixtures): $\hat{p}_i^* = \left(p_i^* - p_{j(C)}\right)/\left(p_{i(C)}^* - p_{j(C)}\right)$. To count the total number of spores produced by a set of fruiting bodies from a given number of cells plated (10^7 cells/plate), we harvested the entire agar discs from the plates into 5mL of spore buffer (20mM EDTA, 0.05% NP40) and counted spores using a hemocytometer.

The ten strains were used to construct 34 different types of chimeric mixtures, with each strain used in at least 4 different pairings. Within each pairing, chimeras were created in which strains were mixed in seven different input frequency combinations (0.05, 0.10, 0.25, 0.50, 0.75, 0.90 and 0.95). For each pair of strains, the set of chimeric mixtures across different input frequencies were independently replicated at least twice (with an average of 4 replicates per pair) for a total of 944 chimeric mixtures composed from the 34 pairs across the various input frequencies. Two of these strain pairings (NC28.1+NC63.2 and NC34.2+NC105.1) were replicated a larger number of times (*N*=18 and *N*=15 replicates respectively) to provide higher

resolution examples. Each experimental replicate therefore provides measurements of the relative representation of each strain in the sporehead and the total number of spores produced by the pair across different input frequencies. Every experimental replicate for a given pair also produced an estimate of the clonal spore production for both strains in the pair.

Estimation of investment and relative fitness

To provide for direct comparison between the model and the experimental data we calculated each parameter from the model following the same methods used to process the data. Four types of measurements were used to generate an estimate of stalk investment of a strain within a chimera ($\hat{I}_{i|\hat{p}_i}$): the total number of spores produced by chimeric fruiting bodies composed from strains *i* and *j* ($\hat{T}_{G(ij)}$), the total number of spores produced by a strain when in clonal fruiting bodies ($\hat{T}_i$), the input proportion of a strain within a chimeric mix ($\hat{p}_i$), and the output proportion of a strain within chimeric sporeheads ($\hat{p}_i'$). From these values we calculated the number of spores from a given strain within the chimeric sporeheads as $\hat{p}_i'\hat{T}_{G(ij)}$. This measure of spore production was normalized against the clonal spore production of the strain to account for any inherent differences in numbers of spores produced by different strains (which reflect differences in spore size and fixed differences in allocation of cells to spores; Buttery *et al.*, 2009; Wolf *et al.*, 2015) to produce a measure of relative spore production: $\hat{T}_{i|\hat{p}_i} = \left(\hat{p}_i'\hat{T}_{G(ij)}\right)/\left(\hat{p}_i\hat{T}_i\right)$. The inverse of the relative allocation of cells to spores provides a measure of relative investment into stalk:

$$\hat{I}_{i|\hat{p}_i} = \hat{T}_{i|\hat{p}_i}{}^{-1} = \frac{\left(\hat{p}_i\hat{T}_i\right)}{\left(\hat{p}_i'\hat{T}_{G(ij)}\right)} \tag{8}$$

Therefore, an investment value ($\hat{I}_{i|\hat{p}_i}$) of 1 indicates that a strain allocates the same proportion of cells to spores when in a chimera as when clonal. Since we expect the allocation pattern of

clones to correspond to the optimal pattern, a value of 1 indicates that cells in both clones and chimeras are allocating a proportion Θ_G of their cells into stalk and $1 - \Theta_G$ into spores. In the case where strains allocate 100% of their cells to spores, the estimate of relative investment ($\hat{I}_{i|\hat{p}_i}$) is expected to simply be the ratio of the clonal level of allocation of cells to spores ($1 - \Theta_G$) to 1 (where 1 is the proportion allocated in a chimera). Thus, an investment value corresponding to $1 - \Theta_G$ is equivalent to a pattern of zero investment of cells into stalk. Therefore, when we present the patterns of investment we rescale the estimates that are based on relative spore production to a scale that reflects relative investment in stalk by simply subtracting a value of $1 - \Theta_G$. As a result, when strains invest at the clonal level we get the expected investment value of Θ_G, and when they allocate all cells to spores (*i.e.*, show zero investment) we get a value of 0. When applying this method to the analysis of data from the natural strains we use an optimal investment value of 30% of cells into the stalk, which is supported by a variety of empirical measurements (Forman & Garrod, 1977; Chattwood *et al.*, 2013). The investment for both strains within each chimeric combination within each experimental replicate were calculated separately.

To calculate relative collective investment for a group ($\hat{I}_G$) we first calculated the number of spores we would expect in a chimera given the clonal spore production for the pair and their relative frequencies in the chimera: $\hat{T}_{G|\text{clonal}(ij)} = \left(\hat{p}_i T_i + \hat{p}_j T_j\right)$. Collective investment was calculated following equation (4) by dividing this clonal expectation by the observed number of spores produced by a chimera:

$$\hat{I}_G = \frac{\hat{T}_{G|\text{clonal}(ij)}}{\hat{T}_{G(ij)}} = \frac{\left(\hat{p}_i T_i + \hat{p}_j T_j\right)}{\hat{T}_{G(ij)}} \tag{9}$$

Collective investment for each chimeric combination was calculated for each experimental replicate using the measures of the component parameters for that replicate. As with the measure of individual investment (eqn. 8), the pattern of collective investment reflects the

relative allocation of cells to spores by strains in a chimera compared to the pattern they shown when clonal (but measured for the entire group, rather than for the individual strains separately). Hence, the values of collective investment calculated using equation (9) have the same scaling as the measure for individual investment (eqn. 8). Therefore, we also subtracted a value of $1 - \Theta_G$ from all collective investment values, such that optimal investment (*i.e.*, the clonal pattern) corresponds to the expected value of Θ_G and the scenario where the collective produces only spores corresponds to a collective investment value of zero.

Relative fitness within a group follows the definition in the model and simply reflects the representation of a strain in the sporehead relative to its input frequency:

$$\hat{\rho}_{i|j} = \hat{p}_i'/\hat{p}_i \quad (10)$$

For simplicity, we compare the fitness of strains using the ratio of their relative fitness values (*e.g.*, $\hat{\rho}_{i|j}/\hat{\rho}_{j|i}$ for strain i relative to j). Values of relative fitness were calculated for each individual replicate. To test for any potential bias caused by the experimental labeling and methods used to calculate relative fitness, we applied the calculation of relative fitness in equation (10) to clonal self-mixes of labeled and unlabeled cells across the same set of frequencies. We find no significant frequency-dependent pattern of relative fitness in these self-mixes ($F_{1, 195} = 1.65$, $p = 0.2$; see SI Appendix, Figure S8).

Patterns of collective investment, individual investment, and relative fitness across frequencies were modelled using a mixed model implemented in SAS (SAS Institute, Cary, NC, USA) fitted by maximum likelihood. For collective investment, frequency was modelled as a quadratic fixed effect with experimental replicate as a random effect. For individual investment and relative fitness, frequency was modelled as a cubic fixed effect. For relative fitness, strain-by-block was included as a random grouping variable, while for investment, strain was included as a grouping variable (owing to a lack of convergence for a model

containing a block or replicate effect). Reduced versions of all models were also run without any fixed effects (*i.e.*, with only the random effects). Significance was determined by calculating twice the difference in the negative log likelihoods of the two models (full model and reduced), which is approximately chi-square distributed with degrees of freedom determined by the difference in the number of parameters in the models.

Measurement and analysis of the cost of chimerism

To measure the risk of fruiting body collapse, we collected two sources of data. First, we created 50:50 chimeric and clonal mixes of ten strain pairs (NC28.1, NC34.2, NC52.3, NC60.1, NC63.2, NC69.1, NC71.1, NC80.1, NC99.1 and NC105.1), with an average of 10.4 replicates per chimeric combination (total N = 469) and 13 replicates per clone (total N = 130) (which together represent data from 31,026 fruiting bodies). Differences between clonal and chimeric mixes were analyzed using a mixed model with aggregation type (clonal or chimeric) as a fixed effect and pair as a random effect. Model degrees of freedom were determined using the Kenward-Roger approximation, which corrects the denominator degrees of freedom for the fixed effect based on the structure of the random effect to avoid pseudoreplication. Second, we created chimeric mixes across a range of focal strain frequencies (0.05, 0.10, 0.25, 0.5, 0.75, 0.90 and 0.95) for six strain pairs (NC28.1+NC105.1, NC99.1+NC105.1, NC99.1+NC60.1, NC34.2.1+NC105.1, NC63.2.1+NC60.1 and NC34.2+NC60.1). Mixes were plated as a 10µl droplet onto non-nutrient KK2 agar in a 24-well dish and allowed to develop into fruiting bodies. The number of fruiting bodies that had spontaneously collapsed was scored as a proportion of the total number of fruiting bodies in the well. Data were modelled using a mixed model implemented in SAS (SAS Institute, Cary, NC, USA) fitted by maximum likelihood with frequency modelled as a fixed quadratic effect and pair as a random grouping variable. Significance was determined by calculating twice the difference in the negative log likelihoods of the two models (see above).

Measurement and analysis of segregation

To measure the degree of segregation between pairs of strains across different asymmetry in relatedness, we followed established protocols for measuring segregation for pairs at equal frequency and applied these methods to measurements across a range of pair-wise frequencies (Ostrowski *et al.*, 2008; Benabentos *et al.*, 2009). Briefly, cells were labelled with CellTracker Green CMFDA (with DMSO used as a control for unlabelled cells) and strains were reciprocally mixed at a range of relative frequencies of the labelled strain (0.05, 0.10, 0.25, 0.5, 0.75, 0.90 and 0.95). Mixes were plated as a 10µl droplet on ~1.25g of sharp horticultural sand (Keith Singleton) wetted with 250µl of KK2 in a 24-well dish and allowed to develop to form fruiting bodies. Individual fruiting bodies were then harvested into spore buffer (KK2 with 20mM EDTA and 0.05% NP40), and the proportion of fluorescent to non-flourescent spores in each fruiting body measured by flow cytometery. We measured patterns of segregation using three different pairs of strains (NC28.1+NC63.2, NC105.1+NC34.2, and NC105.1+NC99.1), with at least 10 sporeheads measured for each pair at each frequency (for a total of 692 sporeheads overall).

A metric of the degree of segregation was calculated following ref. (Gruenheit *et al.*, 2017). Briefly, this measure is based on the standard deviation of a strain's proportional representation across sporeheads ($std(\hat{p}_i')$) at a given input frequency. If there is no segregation, then we would expect all variation in the representation of a strain across fruiting bodies (composed from the same proportions of strains) to be due to random binomial sampling error, and hence $std(\hat{p}_i')$ should be very small given the number of spores counted. However, when there is segregation, we expect to see much more variation in the representation of a strain across fruiting bodies as strains preferentially aggregate with themselves. Because the maximum value of this standard deviation depends on the relative frequencies of the strains, it is standardized to the maximum possible value, which is

determined by the geometric mean of the average representation of the two strains across all sporeheads ($\bar{\hat{p}}_i'$), which is $\sqrt{\bar{\hat{p}}_i'(1-\bar{\hat{p}}_i')}$. This yields a standardized measure of segregation:

$$Segregation_{i,j} = \frac{std(\hat{p}_i')}{\sqrt{\bar{\hat{p}}_i'(1-\bar{\hat{p}}_i')}} \quad (11)$$

which goes from 0 (no segregation) to 1 (the maximum possible degree of segregation, which would necessarily correspond to all fruiting bodies being clonal, with the relative frequency of each type of clonal fruiting body depending on the relative frequencies of the strains). In the statistical analysis, segregation data were modelled using a quadratic model following the approach outlined above for fruiting body collapse.

Acknowledgements

This work was funded by a grant from the Biotechnology and Biological Sciences Research Council (BBSRC) to J.B.W. and C.R.L.T. (BB/M01035X/1; BB/M007146/1), a Wellcome Trust Investigator Award to C.R.L.T (WT095643AIA), a BBSRC studentship for P.G.M., a NERC studentship to L.J.B., and support from the University of Bath Alumni Fund. We thank Matthew Cobb and Christopher Knight, for comments on earlier versions of this manuscript, and two anonymous referees for comments that greatly improved the presentation of our work.

References

Axelrod, R. & Hamilton, W.D. 1981. The evolution of cooperation. *Science (80-.)*. **211**: 1390–1396.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E.A., et al. 2009. Polymorphic members of the lag gene family mediate kin discrimination in Dictyostelium. *Curr. Biol.* **19**: 567–572.

Bourke, A.F.G. 2011. *Principles of Social Evolution*. Oxford University Press.

Bruce, J.B., Cooper, G.A., Chabas, H., West, S.A. & Griffin, A.S. 2017. Cheating and resistance to cheating in natural populations of the bacterium Pseudomonas fluorescens. *Evolution (N. Y).* **71**: 2484–2495.

Bshary, R., Grutter, A.S., Willener, A.S.T. & Leimar, O. 2008. Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature* **455**: 964–966.

Buttery, N.J., Rozen, D.E., Wolf, J.B. & Thompson, C.R.L. 2009. Quantification of social behavior in D. discoideum reveals complex fixed and facultative strategies. *Curr. Biol.* **19**: 1373–1377.

Charnov, E.L. 1977. An elementary treatment of the genetical theory of kin-selection. *J. Theor. Biol.* **66**: 541–550.

Chattwood, A., Nagayama, K., Bolourani, P., Harkin, L., Kamjoo, M., Weeks, G., et al. 2013. Developmental lineage priming in Dictyostelium by heterogeneous Ras activation. *Elife* **2**: e01067

Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature* **373**: 209–216.

Cosmides, L.M. & Tooby, J. 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**: 83–129.

Diggle, S.P., Griffin, A.S., Campbell, G.S. & West, S.A. 2007. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**: 411–414.

Dionisio, F. & Gordo, I. 2006. The tragedy of the commons, the public good dilemma and the meaning of rivalry and excludability in evolutionary biology. *Evol. Ecol. Res.* **8**: 321–332.

Doebeli, M. & Hauert, C. 2005. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* **8**: 748–766.

Doebeli, M., Hauert, C. & Killingback, T. 2004. The Evolutionary Origin of Cooperators and Defectors. *Science (80-.).* **306**: 859–862.

Dugatkin, L.A. & Reeve, H.K. 1998. *Game Theory and Animal Behavior*. Oxford University Press.

Forman, D. & Garrod, D.R. 1977. Pattern formation in Dictyostelium discoideum I. Development of prespore cells and its relationship to the pattern of the fruiting body. *J. Embryol. Exp. Morph.* **40**: 215–228.

Foster, K.R. 2004. Diminishing returns in social evolution: The not-so-tragic commons. *J. Evol. Biol.* **17**: 1058–1072.

Foster, K.R., Fortunato, A., Strassmann, J.E. & Queller, D.C. 2002. The costs and benefits of being a chimera. *Proc. Biol. Sci.* **269**: 2357–2362.

Foster, K.R., Shaulsky, G., Strassmann, J.E., Queller, D.C. & Thompson, C.R. 2004. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**: 693–696.

Frank, S.A. 2010. A general model of the public goods dilemma. *J. Evol. Biol.* **23**: 1245–1250.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press.

Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520–522.

Frank, S.A. 2003. Repression of competition and the evolution of cooperation. *Evolution (N. Y).* **57**: 693–705.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E. & Queller, D.C. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc. Natl. Acad. Sci.* **104**: 8913–8917.

Gilbert, O.M., Strassmann, J.E. & Queller, D.C. 2012. High relatedness in a social amoeba: the role of kin-discriminatory segregation. *Proc. R. Soc. B Biol. Sci.* **279**: 2619–2624.

Gordon, H. 1954. The Economic Theory of a Common-Property Resource: The Fishery. *J. Polit. Econ.* **62**: 124–142.

Gore, J., Youk, H. & van Oudenaarden, A. 2009. Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**: 253–6.

Grafen, A. 2006. Optimization of inclusive fitness. *J. Theor. Biol*. **238**: 541–563.

Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024–1027.

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J., Wolf, J. & Thomposon, C. 2017. A polychromatic "greenbeard" locus determines patterns of cooperation in a social amoeba. *Nat. Commun.* **14171**: 1–9.

Hamilton, W.D. 1963. The Evolution of Altruistic Behavior. *Am. Nat.* **97**: 354–356.

Hamilton, W.D. 1964a. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**: 1–16.

Hamilton, W.D. 1964b. The genetical evolution of social behaviour II. *J. Theor. Biol.* **7**: 17–52.

Hardin, G. 1968. The tragedy of the commons. *Science (80-.).* **162**: 1243–1248.

Ho, H.-I. & Shaulsky, G. 2015. Temporal regulation of kin recognition maintains recognition-cue diversity and suppresses cheating. *Nat. Commun.* **6**: 7144.

Kessin, R.H. 2001. *Dictyostelium: evolution, cell biology, and the development of multicellularity*. Cambridge University Press.

Manhes, P. & Velicer, G.J. 2011. Experimental evolution of selfish policing in social bacteria. *Proc. Natl. Acad. Sci.* **108**: 8357–8362.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press.

Maynard Smith, J. 1974. The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **47**: 209–221.

Maynard Smith, J. & Price, G.R. 1973. Logic of animal conflict. *Nature* **246**: 15–18.

Olson, M. 1965. *The Logic of Collective Action*. Harvard University Press.

Ostrom, E. 1990. *Governing the Commons: the evolution of institutions for collective action.* Cambridge University Press.

Ostrowski, E.A., Katoh, M., Shaulsky, G., Queller, D.C. & Strassmann, J.E. 2008. Kin discrimination increases with genetic distance in a social amoeba. *PLoS Biol.* **6**: 2376–2382.

Parker, G.A. & Maynard Smith, J. 1990. Optimality theory in evolutionary biology. *Nature* **348**: 27-33

Parkinson, K., Buttery, N.J., Wolf, J.B. & Thompson, C.R.L. 2011. A simple mechanism for complex social behavior. *PLoS Biol.* **9**: e1001039.

Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.

Pollak, S., Omer-Bendori, S., Even-Tov, E., Lipsman, V., Bareia, T., Ben-Zion, I., et al. 2016. Facultative cheating supports the coexistence of diverse quorum-sensing alleles. *Proc. Natl. Acad. Sci.* **113**: 2152–2157.

Queller, D.C., Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F. & Smalley, R.E. 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**: 366–367.

Rankin, D.J., Bargum, K. & Kokko, H. 2007. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**: 643–651.

Santorelli, L., Kuspa, A., Shaulsky, G., Queller, D.C. & Strassmann, J.E. 2013. A new social gene in Dictyostelium discoideum, chtB. *BMC Evol. Biol.* **13**: 4.

Shaulsky, G. & Kessin, R.H. 2007. The Cold War of the social amoebae. *Curr. Biol.* **17**: 684–692.

Sinervo, B. & Lively, C.M. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* **380**: 240–243.

Smith, J., Queller, D.C. & Strassmann, J.E. 2014. Fruiting bodies of the social amoeba Dictyostelium discoideum increase spore transport by Drosophila. *BMC Evol. Biol.* **14**: 105.

Strassmann, J.E., Gilbert, O.M. & Queller, D.C. 2011. Kin discrimination and cooperation in microbes. *Annu. Rev. Microbiol.* **65**: 349–367.

Strassmann, J.E. & Queller, D.C. 2011. Evolution of cooperation and control of cheating in a social microbe. *Proc. Natl. Acad. Sci. USA* **108**: 10855–10862.

Strassmann, J.E., Zhu, Y. & Queller, D.C. 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum. *Nature* **408**: 965–967.

Taylor, P.D. 1992. Altruism in viscous populations - an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.

Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.

Taylor, P.D., Wild, G. & Gardner, A. 2007. Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* **20**: 301–309.

Travisano, M. & Velicer, G.J. 2004. Strategies of microbial cheater control. *Trends Microbiol.* **12**: 72–78.

Turner, P.E. & Chao, L. 1999. Prisoner's dilemma in an RNA virus. *Nature* **398**: 441–443.

West, S.A. & Buckling, A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc. Biol. Sci.* **270**: 37–44.

West, S.A. & Gardner, A. 2013. Adaptation and inclusive fitness. *Curr. Biol.* **23**: R577-584.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: R661–R672.

Williams, G.C. 1966. *Adaptation and Natural Selection: a critique of some evolutionary thought*. Princeton University Press.

Wolf, J.B., Howie, J.A., Parkinson, K., Gruenheit, N., Melo, D., Rozen, D., et al. 2015. Fitness trade-offs result in the illusion of social success. *Curr. Biol.* **25**: 1086–1090.

1 Xavier, J.B., Kim, W. & Foster, K.R. 2011. A molecular mechanism that stabilizes cooperative
2 secretions in Pseudomonas aeruginosa. *Mol. Microbiol.* **79**: 166–179.

3

Comment 1: Self-interest and the snowdrift

The preceding paper made claims about the links between the observed behaviour and other general games that have been used to dissect the evolution of cooperation. Here I examine those links in greater detail.

The Collective Investment game is a simple framework for analysing the strategic dilemma of cooperation within a group of players. The central idea that the Collective Investment game brings to the fore is how any social situation can provide opportunities for both cooperation and exploitation. Although the model was phrased with respect to the social amoeba (*Dictyostelium discoideum*), the model setup is generalisable to a wide diversity of different social systems (see Paper 3) because it is free from specific constraints. Further, as players can have quantitatively different frequencies within the group and continuously-variable strategies, players have both the ability to possess different quantitative degrees of incentive to cooperate and the ability to express those preferences in their strategy. Consequently, the scenario is highly revealing: in sweeping terms, the model suggests that a player's ESS involves both increasing levels of cooperation as a player has higher frequency (or increasing levels of defection as a player has lower frequency), and indeed individual strains of the social amoeba are shown to exhibit the ESS pattern of investment into stalk cells (which is a form of cooperation in producing a public good). Thus, players exhibit exploitation when rare and cooperation when common.

What do cooperation and exploitation mean in terms of inclusive fitness? The Collective Investment game does not explicitly use inclusive fitness theory, although it can gel with it (see Paper 3). The stalk-producing behaviour of *D. discoideum* is widely considered to be a paradigmatic case of altruism (*e.g.* Strassmann *et al.*, 2000), which requires considering the actor as the individual cell and the recipients as the spore cells. Whilst this may seem reasonable, this would imply that a cell that differentiates to form a spore cell is a cheater, which seems counterintuitive because, in these terms, only cheaters can possibly survive into

the next generation. This confusion could presumably be avoided by treating the differentiated cell type (of stalk or spore cell) as an incomplete expression of the probabilistic phenotype that is the cell's stalk allocation strategy, wherein a particular strategy appears more like mutual benefit than altruism with reference to what other cells of the same strain do (*i.e.* with the same genotype). Consequently, in Paper 1, we opted for the language that describes strains of *D. discoideum* as the players of the Collective Investment game, which would imply that cooperatively allocating some cells to the stalk is mutually beneficial for the focal strain and their competing non-focal strain. Consequently, not allocating cells to the stalk is exploitative, which is analogous to cheating within a particular social situation, though we would consider a cheater strain as one that never produces any stalk cells. In this way, conditional cheating is possible against a mutually beneficial behaviour because frequency asymmetry which means that the costs and benefits (of inclusive fitness) are not the same for different strains, whilst obligate cheater strains would have lower overall fitness than a conditionally cooperative strain. Thus, the focus on strains not cells helps to focus attention on the differences between strains that we are most interested in (namely, their frequency and the arising conditional cooperation). In this way, Paper 1 would treat the strain as the actor (and the spores as the recipients). This phrasing seems reasonable when considering that all members of a strain share the same genotype. As a result, the language of Paper 1 is closer to a gene's eye view rather than an inclusive fitness perspective, which we find more intuitive in this instance.

Across the different frequency scenarios, the Collective Investment game has similarity to classic bimatrix games, which have been the preoccupation of much of evolutionary game theory (Maynard Smith & Price, 1973; Maynard Smith, 1982; Sigmund & Nowak, 1986; Skyrms, 1996, 2004; Binmore, 2005; Doebeli & Hauert, 2005; Nowak, 2006; Sigmund, 2010). Bimatrix games involve two players with two different possible strategies (cooperate/defect). In Paper 1, similarities are drawn to the prisoner's dilemma when players are at nearly-equal frequency and to the snowdrift when players have more unequal frequencies. In the prisoner's dilemma (Table 1B), the ESS is for players to defect such that

there is no cooperation between the players because there is a temptation to defect irrespective of whether the other player cooperates or not (B>A; D>C). When players are at nearly-equal frequencies in the Collective Investment game, the payoffs of cooperation conform to a prisoner's dilemma – as scaled between maximum and minimum level of cooperation. When one player has high frequency in the Collective Investment game, cooperation can lead to higher fitness than defection even if the other player defects. In this way, the payoffs in the Collective Investment game can conform to another well-studied bimatrix game – the snowdrift. The snowdrift differs from the prisoner's dilemma in the order of its payoffs (see Table 1B, 1C), suggesting that there is still a temptation to defect (B>A) but that it is better to cooperate against a defector than otherwise (C>D). The difference between the Collective Investment game and the snowdrift is that, for the snowdrift, an arbitrary asymmetry, such as flipping a fair coin, could decide which player cooperates and which defects. Whereas, in the Collective Investment game, frequency is not an arbitrary variable, but favours a more common player to take on the burden of cooperation. Further, in the Collective Investment game, there is a penalty for a group of players investing below the optimal level of collective investment so, unlike the snowdrift (but like the prisoner's dilemma), there is a group-level cost from one player defecting.

A simple rule can be used to describe the conditions when a game is more snowdrift-like or prisoner's dilemma-like. A game is snowdrift-like when one player invests into public goods whilst the other player does not, and prisoner's dilemma-like when both players invest at a low rate (see Figure 1). A 'snowdrift rule' can describe when (in a group of players pursuing the ESS) a focal player has an incentive to invest whilst the other player does not. The snowdrift rule can take two specific forms, using the simpler notation from Paper 2: a focal player's frequency/relatedness (r_i), the costs of public goods (c) and the benefits of public goods (b). When the other player does not have any incentive to invest (*i.e.* their ESS level of investment is below zero when the other player does not invest), the snowdrift rule simply requires the focal player to have an incentive to invest (*i.e.* the focal player's ESS is

greater than zero): $r_i > c/b$. When the other player has an incentive to invest, the snowdrift rule requires the focal player to have more of an incentive (*i.e.* the non-focal player's ESS is exactly zero): $r_i > (2b - c)/3b$. As these two limits capture the same snowdrift logic (where only the focal player invests), both rules can be placed into a general form using a new term for the opportunity cost of investment ($k = cr_i x_i = (r_i b - c)/2b$, given $k > 0$), which is the reduction in the share of the benefits that a player suffers from investment:

$$r_i - k > r_{-i} \quad (1)$$

The LHS represents the fraction of the benefits of public goods reaped by the focal player after contributing toward public goods (*i.e.* the number of spore cells in the social amoeba system), whilst the RHS represents the fraction of the benefits of public goods reaped by the other player given that they do not contribute toward public goods. Thus, the snowdrift rule makes explicit that a focal player only invests into public goods, in spite of the other player not investing, when that focal player reaps the majority (>50%) of the benefit from its investment. The two specific forms are reached by the focal player investing ($k > 0$) being more stringent or the player reaping the majority of the benefits ($r_i - k > r_{-i}$) being more stringent. In this way, irrespective of whether or not the social behaviour is viewed as altruism or mutual benefit in an inclusive fitness framework, the snowdrift-like cooperation in the Collective Investment game is thoroughly self-interested, in balancing the cost to self against the benefit to self after dilution of benefits to some nonself (which is because the inclusive fitness framework does not explicitly focus on explaining the quantity of investment – but rather on explaining whether or not a gene for a social behaviour is qualitatively favoured; see Paper 3). The snowdrift rule is likely to be a general rule for mutual benefit in whole-group traits among competitors, with equivalence to mutual benefit under negative relatedness for other-group traits (as described by Hamilton's rule when $c > 0$ and $rb < 0$; see Pepper, 2000), simply stating that the gene for cooperation must reap a larger benefit that its competitor.

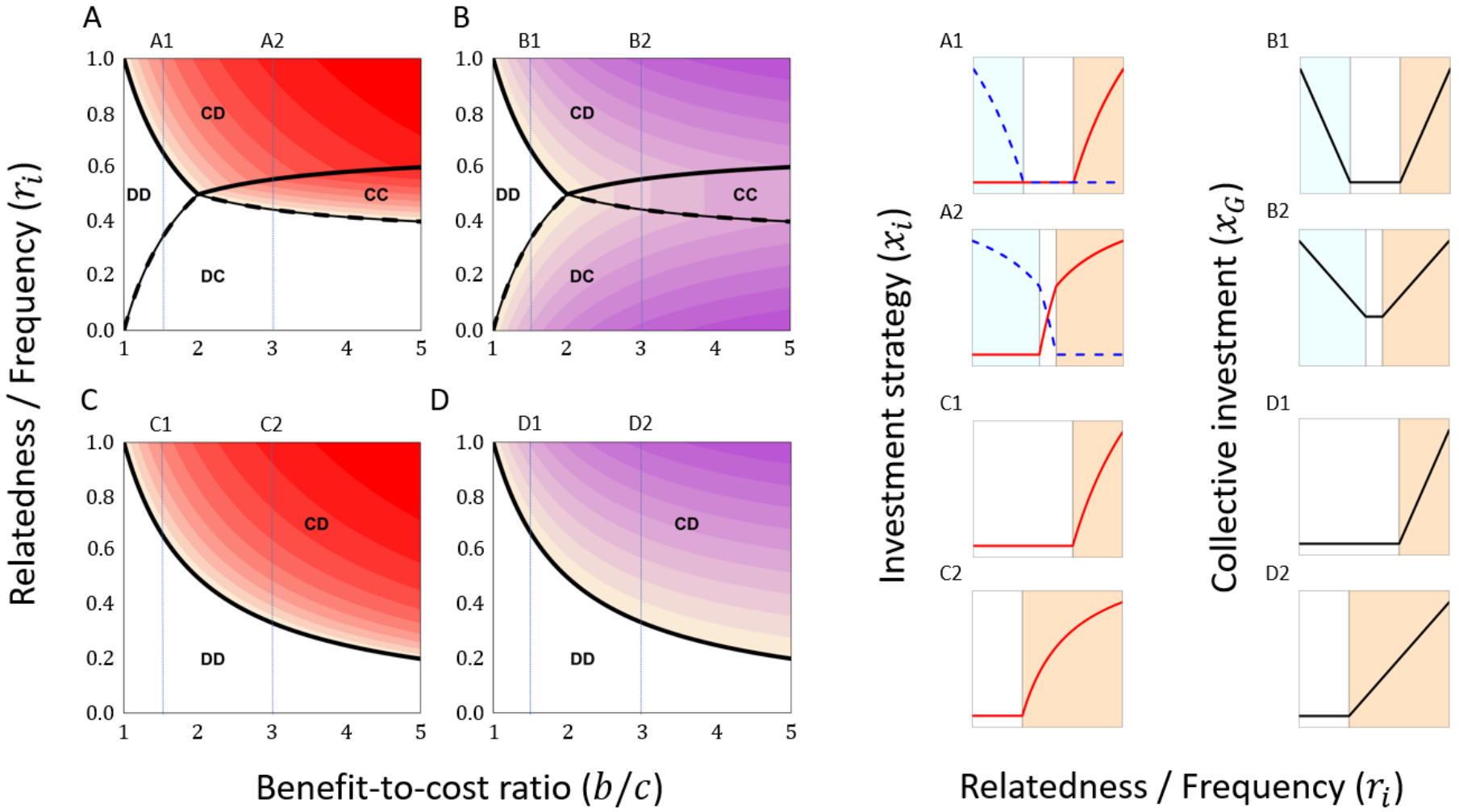

A
A1
A2
B
B1
B2
C
C1
C2
D
D1
D2
CD
DD
CC
DC
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
Relatedness / Frequency (r_i)
Benefit-to-cost ratio (b/c)
Investment strategy (x_i)
Collective investment (x_G)
Relatedness / Frequency (r_i)

Figure 1. The investment (x_i or x_G) across a focal player's relatedness/frequency (r_i) and across benefit-to-cost ratios (b/c; assuming $b > c$ and $c > 1$ so there is a positive and intermediate level of group-optimal investment $0 \leq \varphi \leq 1$). The game-space where both players exhibit the ESS ($x_i = \hat{x}_i$ and $x_{-i} = \hat{x}_{-i}$; **A, B**) and one player plays the selfish investment strategy whilst the other does not invest ($x_i = \tilde{x}_i$ and $x_{-i} = 0$; **C, D**) for investment strategy (x_i; **A, C**) and collective investment ($x_G = r_i x_i + r_i x_{-i}$; **B, D**). The limit for one player providing all the investment for both the focal and non-focal player is marked by a black-line (solid for player one and dotted for player two respectively), dividing the game-space into regions of resource-holding where neither (DD = defect-defect), one (CD/DC = cooperate-defect or visa versa), and both players (CC = cooperate-cooperate) make a contribution to collective investment. Across all panels (**A-D**), to showcase the difference between the investment strategy for individuals and groups across variation in the benefit-to-cost ratio of public goods (b/c), two scenarios for investment strategy at intermediate resource-holdings are presented (labelled #1 and #2 for $b/c = 3/2$ and $b/c = 3$ respectively, where # is any of **A-D**) by extracting a cross-section of the panels **A-D**. The cross-sections are plotted as investment strategy (x_i) across a focal player's resource-holdings (r_i). In each case, the red line represents the focal player with red shading for resource-holdings where the focal player is the sole investor. Additionally, with two-players, the blue line represents the non-focal player with blue shading for resource-holdings where the non-focal player is the sole investor. (NB: the non-focal player's curve is plotted with respect to the resource-holding refers to the focal player's resource-holding which is one minus the non-focal player's resource-holding.)

The nature of the ESS draws parallels with another simple bimatrix game (Table 1D). The stag hunt emphasises that cooperation can be nothing more than a convention, wherein some conventions are better than others. In the stag hunt, as along as both players play the same strategy they reap some benefit, but if players do not then they each receive no benefit. Thus, the evolutionary problem of cooperation in the stag hunt is one of coordination, suggesting that there are gains from both pursuing the same convention. However, as some conventions are better for everyone than others (A>D), there is a problem arising from being unable to obtain the best convention because of the starting point of an evolutionary process (as A,D>B,C) rather than there being some perennial temptation to defect (unlike the prisoner's dilemma or snowdrift). The Collective Investment game has elements of a coordination problem because (especially when the snowdrift rule is not met for either player) the selected level of investment depends on the best strategy in response to the current level of investment by the other player. So, if the other player did not play the ESS, the best response of a player's investment to maximize their fitness would be different. However, as strategies coevolve en route to the ESS, the convention that is arrived at in the ESS reflects the compromise between the different incentives for players to invest at their given frequencies – thereby resolving any 'stalemate' arising from being stuck with one convention or other. Further, even when players have the same frequency, the ESS comes to reflect this compromise because any investment has a two-fold cost: a cell can either differentiate to become a stalk or spore cell, and so an increase in investment involves both taking away one spore cell and adding one stalk cell (creating a difference of two between the numbers of spore and stalk cells). Therefore, although the ESS for the Collective Investment game has similar properties to a coordination problem, the evolutionary path to the ESS avoids the stalemate of the stag hunt because the biology ensures that strategy coevolution is convergent.

Paper 1 has established that strains of *D. discoideum* behave as conditional cooperators, and the additional discussion in Comment 1 has expanded on how the game changes its nature across different frequencies with appeal to other well-known games. Next,

I move to consider testing the predictions of the Collective Investment game when there are more than two strains of *D. discoideum*, which is revealing about the mechanism through which conditional cooperation is made possible.

Table 1. Three classic bimatrix games (of two players with two strategies) for understanding the evolution of cooperation. **A)** The general form of a bimatrix game for cooperation, showing symmetry between players' payoffs. **B)** The prisoner's dilemma: B>A>D>C (Flood, 1952; Poundstone, 1993). **C)** The snowdrift: B>A>C>D (Rapoport & Chammah, 1966; Sugden, 1986). **D)** The stag hunt: A>D>B=C (Lewis, 1969).

A) Bimatrix Game	**Cooperate**	**Defect**
Cooperate	A, A	B, C
Defect	C, B	D, D

B) Prisoner's Dilemma	Cooperate	Defect
Cooperate	2, 2	-1, 3
Defect	3, -1	0, 0

C) Snowdrift	**Cooperate**	**Defect**
Cooperate	2, 2	1, 3
Defect	3, 1	0, 0

D) Stag Hunt	**Cooperate**	**Defect**
Cooperate	2, 2	0, 0
Defect	0, 0	1, 1

Paper 2

This declaration concerns the article entitled:			
The not-so-tragic commons in a social microbe			
Publication status (tick one) **Draft manuscript** ✓ **Submitted** ☐ **In review** ☐ **Accepted** ☐ **Published** ☐			
Publication details (reference)	Belcher, L. J., **P. G. Madgwick**, C. R. L. Thompson, and J. B. Wolf. n.d. The not-so-tragic commons in a social microbe.		
Copyright status (tick the appropriate statement) I hold the copyright for this material ✓ Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Formulation of ideas: 30%: PGM contributed to the research concept. Design of methodology: Experiment: 0% Model 40%: PGM significantly contributed to model design, analysis and interpretation of data. Experimental work: 0% Presentation of data in journal format: 15%: PGM contributed to revising and editing the manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	PGMadgwick	**Date**	07/11/2019

Data access statement: email lb780@bath.ac.uk

The not-so-tragic commons in a microbe

Authors: Laurence J. Belcher[1*], Philip G. Madgwick[1], Christopher R.L. Thompson[2], and Jason B. Wolf[1]

Affiliations:

[1] Milner Centre for Evolution and Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, BA2 7AY, UK

[2] Centre for Life's Origins and Evolution, Department of Genetics, Evolution and Environment, University College London, Darwin Building, Gower Street, London, WC1E 6BT, UK

[*] Correspondence to: lb780@bath.ac.uk

Abstract

Individuals across the tree of life make costly contributions that benefit their group. Such cooperation through 'public goods' presents a dilemma as all individuals can benefit from being 'selfish' by withholding contributions whilst benefitting from the contributions of others – leading to the breakdown of cooperation known as the 'Tragedy of the Commons'. Whilst the threat of the tragedy is well understood, studies are largely limited to qualitative descriptions of how 'cooperators' adaptively avoid exploitative 'cheaters', rather than on how all individuals face strategic trade-offs governing how much they should contribute to public goods. Here, we address these limitations with a combination of quantitative theoretical predictions and experimental tests in natural strains of *Dictyostelium discoideum*. Our model broadly predicts that strains invest more in the group when their relatedness to the group is high, as their fitness interests are closer aligned with that of the group, suggesting that groups with few strains likely avoid the tragedy due to enough strains having sufficient motivation to contribute to the group. However, the model also predicts that groups with low average

relatedness should show catastrophic failure due to insufficient contributions. Experimental measures support these predictions, with strains reducing contributions as a function of relatedness to the group. However, despite a close quantitative match between predicted and measured contributions, we surprisingly find that strains avoid the worst of the tragedy. This is mostly likely due to non-adaptive constraints in information and strategy preventing strains from fully withholding contributions and expressing their selfishness. By viewing the tragedy of the commons as the consequence of the rational behaviour of all individuals, rather than 'exploitation' by rouge cheaters, we highlight how non-adaptive constraints on strategy and information can play important and underestimated roles in avoiding the worst of the tragedy.

Keywords: cooperation; conflict; public goods; social behaviour; kin selection

Introduction

Individuals often act in ways that benefit their group (West *et al.*, 2007a; Bourke, 2011). These cooperative acts are typically costly to the actor, yet the benefits are often available to all, regardless of individual contributions. Such cooperation through 'public goods' is therefore vulnerable to selfish individuals who restrict their contributions to public goods, while reaping the benefits from contributions made by others (Olson, 1965; Rankin *et al.*, 2007). This lack of motivation to contribute to public goods underlies the 'Tragedy of the Commons' (Hardin, 1968), where selfish behaviours that maximise personal interests lead to suboptimal group success (Olson, 1965). The potential risk from such tragedies is well-documented in both economics (Gordon, 1954; Ostrom, 1990) and biology (Wenseleers & Ratnieks, 2004; Rankin *et al.*, 2007; Strassmann & Queller, 2014). Furthermore, the persistent threat of the tragedy of the commons can play a critical role in shaping the evolution of cooperation and group organization in nature (Rankin *et al.*, 2007; Frank, 2009); cooperation can be lost when the costs imposed by exploitation outweigh the potential rewards from

contributing, group membership can be restricted to avert the most tragic outcomes, or groups can collapse entirely. Understanding 'why' and 'how much' the tragedy occurs (and how it can be avoided) is therefore of great importance in understanding the evolution of the diversity of cooperative behaviour found in nature.

Our current understanding of the properties of the tragedy in natural systems is largely limited to qualitative descriptions; whether the outcome is catastrophic where no individuals contribute to cooperation, or merely sub-optimal, where contributions to cooperation are simply lower that the level that maximises group success (Wenseleers & Ratnieks, 2004; Rankin *et al.*, 2007; Frank, 2009). Consequently, whilst we have some understanding of the broad-scale, qualitative conditions that lead to 'tragic' outcomes for groups, we have limited understanding of the processes governing the finer-scale variation in public goods production that is important for the evolution of social traits. Whilst there has been discussion of how the tragedy can be avoided (Frank, 1995, 2009; West *et al.*, 2002; Foster *et al.*, 2004, 2006; Rankin *et al.*, 2007), the focus is often on how groups avoid obligate 'cheaters' who impose a 'cheater-load' (Velicer, 2003; Travisano & Velicer, 2004; Van Dyken *et al.*, 2011) on group fitness. Such a perspective masks the fact that contributing to a public good is a strategic choice for all individuals, subject to trade-offs between costs and benefits. Furthermore, in many public goods strategies are not limited to discrete 'cooperate' or 'cheat' strategies. Instead, selection can favour continuous strategies that adaptively adjust contributions to public goods in different social contexts. Studies in microbes have indicated that even simple organisms can strategically adjust their contributions to public goods (Madgwick *et al.*, 2018), suggesting that the manifestation of the tragedy of the commons in nature might be quantitative and depend on social context. Therefore, to understand to extent to which individuals in groups will suffer from the tragedy of the commons and what factors allow groups to collectively avoid the worst of these conditions, we need to consider the strategic selfish behaviour of all members, not just how the group resists those who cheat by not contributing their 'fair share'. To achieve this goal, we need to connect empirical studies to a theoretical framework that

considers how and why individuals strategically vary their contributions to public goods in response to social setting.

The most obvious feature of groups that will drive variation in the contributions individuals make to public goods is their relatedness to group members. Individuals should presumably be more willing to contribute to public goods when benefits go to relatives. In this way, relatedness aligns the fitness interests of an individual with those of the group, because genes in the individual ultimately benefit when the group as a whole prospers (Foster *et al.*, 2006; Taylor *et al.*, 2007; West *et al.*, 2007a; Frank, 2009; Gardner & West, 2014). Hence, individuals should contribute to public goods dependent upon their relatedness to their group – a logic captured by Hamilton's rule ($rb - c > 0$, where b and c capture the benefits to recipients and costs to actors of some cooperative act and r captures the relatedness of the actor to the recipients) (Hamilton, 1964b; Charnov, 1977). The importance of this economic balance for the evolution of cooperative strategies can be overlooked by a perspective that focuses on cheating.

Here, we apply a model of quantitative variation in 'investment' (*i.e.* contribution) into public goods (Madgwick *et al.*, 2018) to experimental analyses of cooperation in the social amoeba *Dictyostelium discoideum*. In this system, individual cells aggregate to collectively form a fruiting body constructed of a stalk (the public good) that facilitates dispersal of spores (the benefit from public goods) (Strassmann *et al.*, 2000; Kessin, 2001; Smith *et al.*, 2014). Because different strains can co-aggregate to form a chimeric fruiting body, the relatedness of a strain to the group can vary (depending on their relative frequency within, or relatedness to, the group), which can shift the balance of costs and benefits determining whether to contribute to public goods. Indeed, previous work examining simple two-strain groups demonstrated that strains can strategically alter their contribution to the stalk in response to relatedness (Madgwick *et al.*, 2018). However, it is unclear how collective contributions to stalk production vary across groups with more complex social compositions, and hence how the tragedy of the commons is manifested in this system – though theoretical

modelling has more generally shown that relatedness has an important role in averting the tragedy of the commons (Frank, 1995, 1996, 2009). Here, we show theoretically that groups containing few strains are likely to have sufficiently high relatedness to the group to maintain a level of stalk production that avoids the collapsing tragedy of the commons (where there are no public goods produced). However, the model also robustly predicts that groups in which relatedness drops below a critical threshold should show catastrophic failure due to a collapsing tragedy of the commons. Experimental measurements of stalk production across a broad range of levels of relatedness support the model predictions, with groups reducing collective contributions to stalk production as relatedness declines in groups, leading to the quantitative manifestation of the tragedy of the commons. Surprisingly, however, despite a close match to model predictions that indicate that groups should show zero contribution to public goods and experience a collapsing tragedy, they avoid the most catastrophic outcomes (where all fruiting bodies fail to support spores). We suggest that this is most likely due to biological constraints in strategies and information that prevents strains from showing zero contribution to the stalk. Therefore, while groups clearly suffer from the tragedy of the commons, the outcomes are not nearly as tragic as predicted by theory because biological constraints prevent the necessary catastrophic decline in contributions to public goods.

Results

Theory predicts groups avoid tragic collapse across a range of relatedness

To make predictions about how the contributions by members of a group contribute to public goods we model cooperation as an evolutionary game, where the players are genotypes (strains) that can make costly quantitative 'investments' (*i.e.* contributions) to public goods, and receive returns on those investments though the benefits to the group as a whole (see Methods). We build on a simple form of the game that was developed to model pairwise interactions between players (Madgwick *et al.*, 2018), and hence cannot be used to

analyse of patterns in more complex groups composed of multiple strains (where the tragedy of the commons is most likely to play out). As expected, the model predicts that motivation to invest into public goods increases with relatedness. More importantly, the model provides clear predictions about specific patterns of investment that are qualitatively robust to a large range of benefits and costs of investment and non-linearity of benefits (see Methods). More specifically, the model predicts that groups with a small number of strains (where at least one strains will have high enough relatedness to the group to motivate investment) will typically make sufficient collective investment to avoid a collapsing tragedy of the commons (where collective investment is at or very close to zero). As such, groups with many strains that all have low relatedness to the group are predicted to show very low investment that reaches zero investment and a catastrophic 'collapsing' tragedy of the commons when relatedness drops below a critical lower threshold.

To facilitate the application of the model to empirical patterns we allow for error in the measurement of relatedness (meaning individuals do not have access to perfect information about their relatedness, rather they estimate their relatedness to the group with some degree of stochastic error or 'noise') Furthermore, we allow error to be frequency-dependant, such that strains make the largest errors in relatedness estimation when at intermediate relative frequency within the group. In this way, specific quantitative patterns of the tragedy of the commons depend on the exact costs (c) and benefits (b) of production of public goods (*i.e.* the stalk), as well as the magnitude of stochastic of error (e), and how frequency-dependant (f) errors in the estimation of relatedness are. Hence, for any given level of relatedness the model provides a predicted level of investment that can be evaluated empirically.

Data best fit a model with high benefits relative to costs and intermediate error

To understand patterns of collective investment and the threat of the tragedy of the commons, we first examine how investment changes across variation in relatedness in groups each composed of three different strains of *D. discoideum* (replicated across several different sets of strains). Relatedness to the group can easily be manipulated in *D. discoideum* because it is equivalent to the frequency of a strain in a group. Because relatedness of three strains can vary across a huge array of possible combinations, we assess patterns of investment by varying relatedness along several 'transects', where we hold the relatedness of one strain constant while varying the relatedness of the others (see Figure 1). We fit these data to the predictions from the model to identify the best fit estimates for the costs (c), benefits (b), and the total level of (e) and degree of frequency-dependence (f) of the error in the strains' estimation of their relatedness to the group. We find that the best fit occurs when the benefits from stalk investment far exceed the costs c = 1 and b = 8) and strains make fairly large errors in measurement of relatedness (e = 0.4) that are moderately frequency-dependant (f = 0.175). Moreover, we find that experimental measures of collective investment in groups of three strains are not significantly different from the values predicted by the model (Paired t-test: $t_{(19)} = -1.463, p = 0.160$) (Figure 1), indicating a close match between the predicted and observed values. The relatively high benefit to cost ratio suggests that individuals will be incentivised to invest into public goods across a large range of relatedness values, while the presence of a relatively large estimated error in measurement of relatedness means that we expect a slower decline in collective investment as relatedness declines than the perfect information model, with a lower threshold below which no strain should invest.

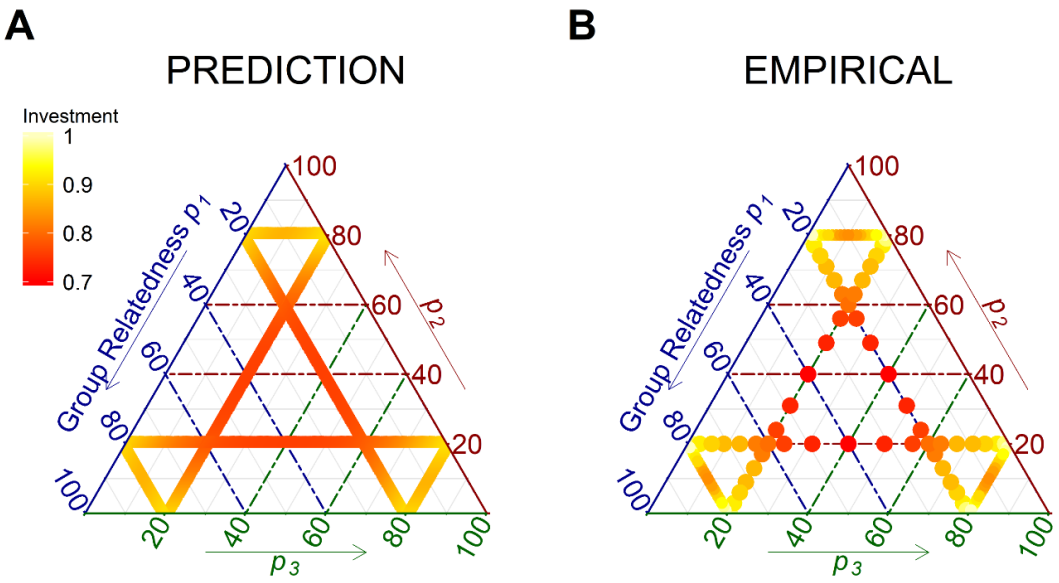


Figure 1. Predicted (**A**) and measured (**B**) Collective Investment in the three-player game for transects through the available game-space of possible group compositions. Two transects are used, where a focal strain has a fixed relatedness to the group (=0.2 or 0.8), and the relative whole-group relatedness of the other two players varies. The colour represents Collective Investment (see scale), with red representing low collective Investment, and yellow high Collective Investment. Predictions are shown for the parameter set that provides the best fit to the data (C=1, B=8, E=0.4, F=0.175).

Theory predicts groups collapse with many strains at low relatedness

The close match between values of collective investment predicted by the model and the experimentally observed values indicates that our model can be utilised to make accurate quantitative predictions of public goods investment across social contexts. Therefore, to understand how and when the pattern of collective investment is expected to lead to the tragedy of the commons, including the potential for the collapsing tragedy, we created an array of groups with declining levels of relatedness. For this we mixed sets of *N* strains in equal proportions, such that relatedness equals 1/*N* for each group (and declines as more strains are added to the groups). The model predicts that these *N*-player groups should show a decline in collective investment down to a relatedness of about 0.1, below which relatedness will be too low to incentivise any level of investment, and hence collective investment should be zero (Figure 2A). In agreement with model predictions, we find that collective investment shows the expected deterioration as relatedness declines (mixed model: $\chi_1^2 = 69.1$, N=228, $p < 10^{-15}$), eventually reaching a plateau at the level of relatedness where collective investment should go to zero (Figure 2A). The observed pattern shows a close fit to that predicted based on the parameter estimates from the three-strain experiments, where the observed values are not significantly different from those predicted by the collective investment game (paired t-test: $t_{(9)} = 0.4034, p = 0.403$). To confirm the agreement between predictions from the three-strain experiments and the observations in the *N*-strain experiment we conducted an independent analysis where we fit the data from the *N*-strain experiments to predictions from the Collective Investment game. Remarkably, we find that that the best fit parameter estimates are identical to those yielded by the model fit to the three-strain data ($c = 1$, $b = 8$, $e = 0.4$, and $f = 0.175$) which strongly supports the inference of the model.

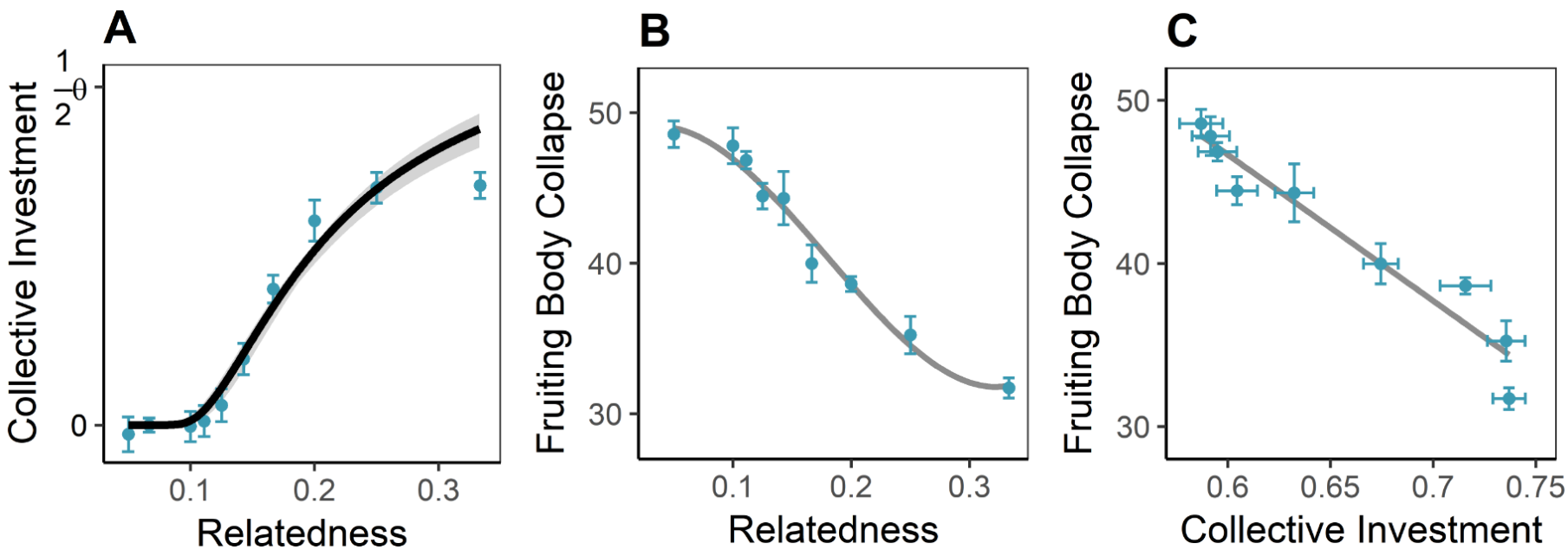

A
B
C
Collective Investment
1
1/2
0
0.1
0.2
0.3
Relatedness
Fruiting Body Collapse
50
40
30
0.1
0.2
0.3
Relatedness
Fruiting Body Collapse
50
40
30
0.6
0.65
0.7
0.75
Collective Investment

Figure 2. **A**) Measured Collective Investment (blue points) and predicted Collective Investment (black line) from the best-fit model for groups varying in size from 3-20 strains with all strains having the same relatedness to the group (=1/N). Grey shading represents confidence intervals around the estimate of the best-fit parameters to the data. **B**) Fruiting body collapse (blue points) for groups varying in size from 3-20 strains with all strains having the same relatedness to the group (=1/N). The grey line shows the fit from a mixed model of fruiting body collapse varying with average relatedness. **C**) Relationship between Collective Investment and fruiting body collapse (blue points). Error bars represent standard errors in Collective Investment (horizontal) and fruiting body collapse (vertical). The grey line shows the correlation between the variables.

Groups suffer when public goods decline, but not as badly as expected

The close fit between model predictions of collective investment and empirical data suggests that groups with low relatedness should experience the catastrophic consequences of the collapsing tragedy of the commons, where there is no stalk produced and groups are unable to take advantage of stalk facilitated dispersal (Figure 2A). However, because our measure of investment is in relation to how strains shift allocation of cells away from stalk and towards production of spores (Madgwick *et al.*, 2018), the pattern represents relative rather than absolute investment. We therefore directly measured the impact of collective investment on groups by evaluating the structural integrity of the stalk produced. To this end, we measured the proportion of fruiting bodies that spontaneously collapsed. We find that the proportion of failed fruiting bodies increases significantly as relatedness declines (mixed model: $\chi_1^2 = 68.0$, N=175, $p < 10^{-15}$) (Figure 2B) and is significantly correlated with the level of collective investment (Pearson correlation: $t_{(9)} = -10.2, p < 10^{-5}, r = -0.97$) (Figure 2C). However, despite the fact that patterns of collective investment closely match those predicted by the model, the groups in which we expect to see zero stalk investment remarkably still built complete fruiting bodies (with stalks), though they suffer from a relatively high rate of fruiting body collapse. This result suggests that, whilst investment is plateauing at low relatedness as predicted, the plateau occurs not at zero investment, but at some limit that potentially reflects a constraint on the available investment strategy that prevents zero investment.

Discussion

Our theoretical analysis of cooperation through production of public goods indicates that groups are expected to make relatively large collective contributions to stalk production across a large range of relatedness values, but these contributions should decline sharply when relatedness within groups falls below a critical threshold. This pattern occurs because strains with low relatedness to a group are necessarily at low frequency in the group and hence their

individual contribution can only have a small impact on the success of the group. For example, even if a strain at low frequency in a group were to make a very large relative contribution to stalk formation (*e.g.* sacrifice half of their cells to produce the stalk), their sacrifice would produce little if any group level benefit because it would represent an insignificant absolute contribution to the stalk. Thus, the potential rewards from contributing simply cannot compensate for the costs. We find a strong quantitative match between model predictions and empirical data for groups containing three strains, with total contribution ('collective investment') to the stalk declining with relatedness because strains lack sufficient motivation to invest (Figure 1). This leads to a robust prediction that contributions to stalk should eventually decline to zero when relatedness falls below a critical lower limit (which generally corresponds to groups containing an increasing number of strains). We find that measured contributions to stalk production matches this prediction (Figure 2A). However, we surprisingly show that even in these most tragic scenarios, groups still produce some stable fruiting bodies (Figure 2B). Whilst groups clearly suffer increasingly from the tragedy of the commons as contributions to stalk production decline, contributions are not falling to zero as predicted by theory. To understand these enigmatic findings, we focus on how groups might avoid the tragedy of the commons, and why the biology of the system may violate the assumptions of the theory in a way that essentially spares groups from the worst of the tragedy.

There are many mechanisms that could lead groups to avoid the worst of the tragedy of the commons, including mechanisms that enforce contributing to public goods (Ågren *et al.*, 2019), such as coercion (Frank, 1995; Wenseleers *et al.*, 2004), rewards (Trivers, 1971; Sasaki & Uchida, 2014), punishment (Gardner & West, 2004a; Bshary & Grutter, 2005; Boyd *et al.*, 2010), and sanctions (Pellmyr & Huth, 1994; Wang & Shaulsky, 2015). However, although such enforcement measures are well-studied across species (Clutton-Brock & Parker, 1995; Frank, 1995, 1996; Rankin *et al.*, 2007), including in microbes (Manhes & Velicer, 2011; Wechsler *et al.*, 2019) they arguably have more limited relevance in microbes due to the presence of clonal growth that structure microbial populations (Nadell *et al.*, 2010; Ågren

et al., 2019) and limited ability for microbes to identify non-cooperators whose contributions must be 'enforced'. Successful cooperation and the avoidance of the tragedy of the commons in microbes is therefore more likely to rely on high relatedness in groups, which restricts within-group competition and aligns the fitness interests of individuals with the group. Such outcomes can be achieved by excluding non-kin, or preferentially interacting with kin (Dionisio & Gordo, 2006, 2007). In *D. discoideum*, strains have been shown to be able to implement a mechanism that allows them to partially segregate away from other strains they encounter in aggregations (Ostrowski *et al.*, 2008; Benabentos *et al.*, 2009; Gruenheit *et al.*, 2017), resulting in preferential interactions with partners who match at a polymorphic recognition locus (Gruenheit *et al.*, 2017). It appears likely that this mechanism evolved as a way to avoid the most tragic conditions (which occur at intermediate levels of relatedness) given that segregation is frequency dependent, with strains segregating most when they are at intermediate relatedness to the group (Madgwick *et al.*, 2018). However, while segregation can help groups avoid the most tragic conditions, it is an imperfect mechanism that does not result in perfectly clonal fruiting bodies but rather, is quantitative and increases the variance in relatedness across fruiting bodies, which thereby inflates the average relatedness within fruiting bodies. Therefore, strains are still likely to experience the full range of relatedness in nature, which leaves groups vulnerable to the tragedy of the commons, even with a mechanism in place to reduce its impact (Gruenheit *et al.*, 2017).

Although much consideration has been given to adaptive mechanisms for avoiding the tragedy, there are many non-adaptive constraints that can preserve cooperation. One such constraint in *D. discoideum* could be an inability to down-regulate stalk cell fate to zero. Strains produce and respond to a various factors that regulate cell fate (Morris *et al.*, 1987), which affect allocation of cells to stalk production (Parkinson *et al.*, 2011). Such factors may present an opportunity for coercion of others to invest in the stalk, and may combine with pleiotropic constraints (*e.g.* Foster et al. 2004) in restricting how low a strain can invest. A further constraint is the inability of strains to perfectly assess their relatedness to the group

(Madgwick *et al.*, 2018), leading a strain to contribute more to the stalk than may be optimal when at low relatedness to the group (SI Appendix, Figure S1A). Whilst a perfectly adapted individual would have information about its relatedness to all social partners, and choose its strategy accordingly (Hamilton, 2001), such information could generate substantial conflict between social partner (Ratnieks & Reeve, 1992). In this way, informational constraints can be an important non-adaptive mechanism for avoiding the tragedy of the commons, benefitting the group by restricting a strain's ability to express its optimal selfishness. Imperfect information of this kind plays an important role in alleviating potential tragedies in many taxa, such as polyandrous social insect workers failing to favour their own patriline (Breed *et al.*, 1994; Keller, 1997; Nonacs, 2011), and cooperatively breeding birds failing to recognise extra-pair young (Dickinson, 2004; Komdeur *et al.*, 2004) despite the obvious inclusive fitness advantage. A lack of information can change the optimal strategy in favour of cooperation, as occurs in meerkats where errors in relatedness estimation can select for indiscriminate altruism (Duncan *et al.*, 2019), and mice where communally nesting mothers contribute milk according to total group size, rather than strategically investing according to their own relatedness (litter size) to the group (Konig, 1994; Ferrari *et al.*, 2015; Ferrari & König, 2017). Overall, the types of constraints on selfishness we observe here, driven by imperfections and constraints in information, strategies, and enforcement measures may be important determinants of the degree to which groups avoid an expected tragedy of the commons throughout nature, particularly in microbes where avoiding such constraints may be more difficult.

Methods

A model of cooperation through public goods

To understand individual and group contributions to public goods we generalise the 'Collective Investment game' (Madgwick *et al.*, 2018), which was originally developed to model pairwise interactions between strains of *Dictyostylium discoideum*. Using the model,

we evaluate how strains should quantitatively adjust their contributions to stalk production as a function of their relatedness to their group. Therefore, the model considers the stalk to be a public good that benefits the group by holding reproduce spores aloft for dispersal (Strassmann *et al.*, 2000; Smith *et al.*, 2014). Groups are composed of N strains that each have a relative frequency of r_i within a group. Here, frequency within a group is equivalent to a strain's whole-group relatedness to the group (*i.e.* relatedness to the group, including self). This allows us to link theoretical predictions (of behaviour with respect to relatedness) directly with empirical results (from manipulations of frequency). To allow a broader discussion of the implications of the model, we use relatedness throughout, whilst noting the equivalence with frequency. Each strain makes an 'investment' by contributing a proportion x_i of its cells into the stalk, with the residual proportion $(1 - x_i)$ going into reproductive spores. Contributions to the stalk (x_i) can therefore vary between 0 and 1 (see Figure 3 for a schematic representation of the model). The benefits of the stalk depend on the total contribute of all strains (*i.e.* the level of 'collective investment') in the group (x_G), which is simply the sum of the proportional contributions of all strains weighted by their frequency in (and hence relatedness to) the group:

$$x_G = \sum_{i=1}^{N} x_i r_i \quad (1)$$

The fitness benefit to the group (B_G) from the stalk is modelled as a linear function in which fitness increases from a baseline value of 1 at a rate of b per unit of group contribution to stalk:

$$B_G = 1 + b x_G \quad (2)$$

Because all members of a group get the same benefit from the stalk regardless contribution, B_G can also be interpreted as the 'between-group' component of fitness. Contributing to the

stalk comes as a personal 'opportunity cost' (see Figure 3) to a strain because the cells it sacrifices to produce the stalk lose out on the opportunity to benefit from the stalk produced. This fitness cost to a strain arising from their contribution to the stalk (C_i) is modelled as a linear function, where fitness declines from a baseline value of 1 by c units for each unit of cells contributed to the stalk (which is a proportion of all of the cells from that strain):

$$C_i = 1 - cx_i \quad (3)$$

Because a strain sacrifices a proportion of its cells (that could have been allocated to spores) to produce the stalk, the fitness cost component (C_i) can also be considered as a strain's 'within-group' component of fitness. The total fitness of a strain (ω_i) is the product of the fitness cost from contributing to the stalk and the fitness benefit from the total stalk produced by the group:

$$\omega_i = C_i B_G = (1 - cx_i)(1 + bx_G) \quad (4)$$

Because a given strain does not have control over the total amount of stalk produced by their group (x_G), the benefit they get from their own contribution to the stalk (B_i) will depend on their frequency in (*i.e.* relatedness to) the group, which determines how much their contribution can impact the total stalk production, *i.e.* $B_i = 1 + br_i x_i$. Therefore, we can evaluate the fitness of a strain in relation to the costs and benefits arising from their personal contribution to stalk production:

$$\widetilde{\omega}_i = C_i B_i = (1 - cx_i)(1 + br_i x_i) \quad (5)$$

which emphasises the fact that the relevant benefit from contributing to the stalk is that which arises from the strain's own contribution (and hence, it is the benefit from their own contribution that incentivises them to contribute).

The optimal proportion of its cells that a strain should contribute ($\hat{x}_i$) to the stalk (in terms of the marginal effect its contribution has on its own fitness) as a function of its relatedness to a group can be solved by finding the level of contribution that maximises fitness using equation (5):

$$\tilde{x}_i = \frac{1}{2}\left(\frac{1}{c} - \frac{1}{br_i}\right) \quad (6)$$

The relationship between equation (6) and Hamilton's rule (Hamilton, 1964b) can be seen by examining the threshold between contributing and not-contributing to stalk formation (*i.e.* the threshold above which $\tilde{x}_i > 0$), which occurs when $r_i b - c > 0$. The patterns of investment predicted by equation (6) are illustrated in Figure 4, which indicates that strains are expected to show zero investment when their relatedness to the group drops below some critical threshold and is expected to approach the level shown by clonal groups (which is taken as the optimal level of investment) as their relatedness approaches 1.

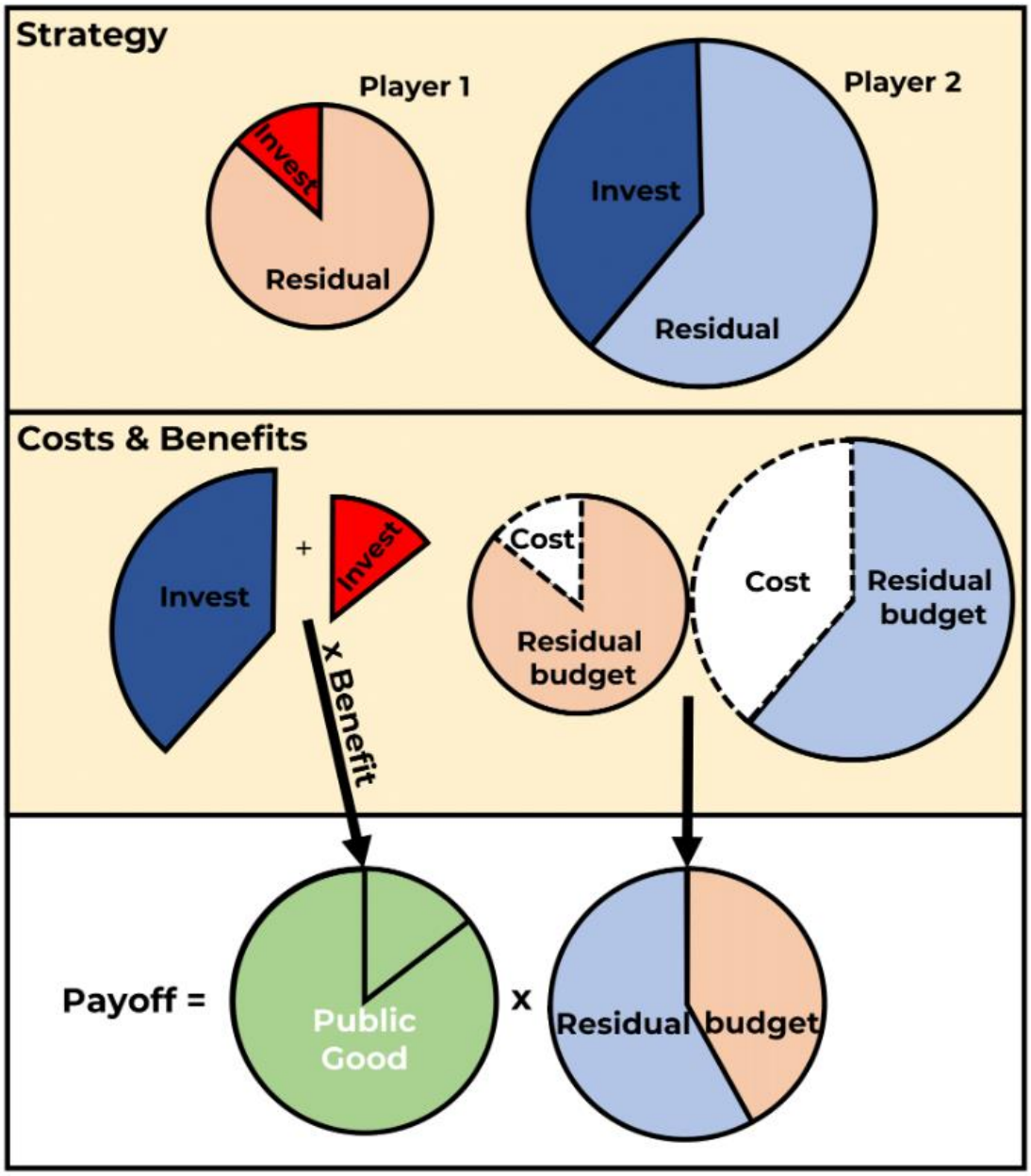


Figure 3. Schematic representation of the 'Collective Investment game' for two strains. Each strain has a total budget (*e.g.* number of cells) based on their relative frequency within the group. Strains invest a proportion of their budget into the public good, with the remainder representing the residual budget which is available to reap benefits. The benefits of investment come through a multiplication *b* of the investments of both players to form the public good. The costs of investment come through a personal 'opportunity cost' of reducing the size of the residual budget of a strain (through which payoffs can be accrued) by investment multiplied by ***c***. The payoff that each player receives is the product of the size of the group's public good

and the relative representation of a player in the fitness accruing cells, which depends on both their relative frequency within the group to start with, and the amount of their residual budget spent on investment.

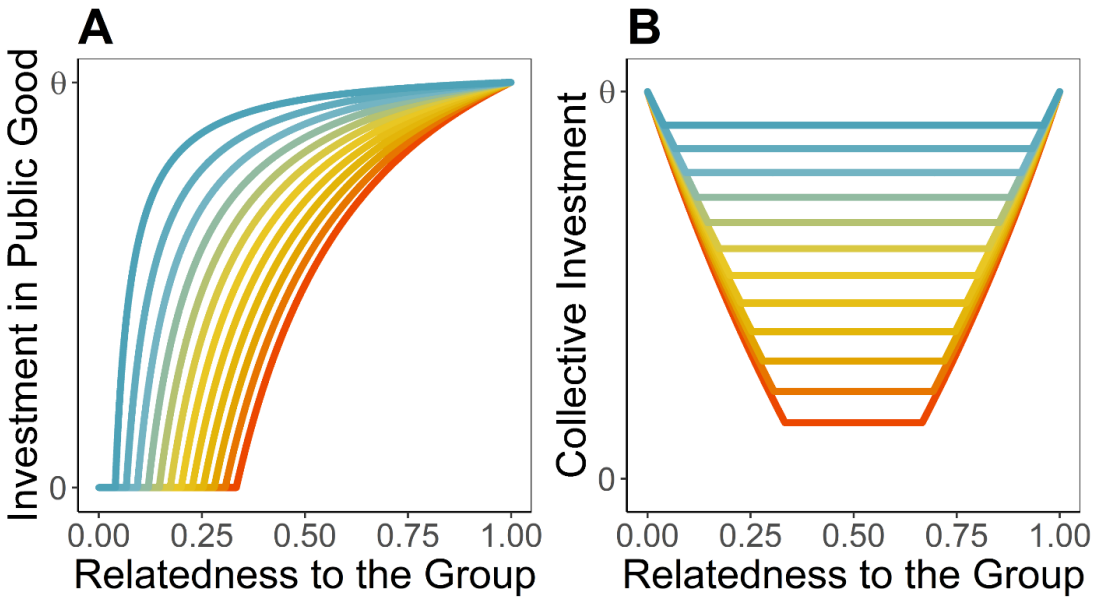


Figure 4. Investment into public goods when the strength of selection on collective investment varies. **A**) Optimal investment for one strain as a function of relatedness to the group. **B**) Collective Investment of two strain as a function of a focal strain's relatedness to the group. In both panels, the red line represents the weakest selection on public goods investment (S=3), and the blue line represents the strongest selection (S=53).

To understand how clonal groups should contribute to the stalk, we can evaluate equation (5) for the case where r_i = 1 (or identify the value x_G of that maximises equation 4), which gives an optimal level of contribution to stalk (θ) of:

$$\theta = \frac{b - c}{2bc} \tag{7}$$

The sum of the contributions made by a group of strains contributing at the level given by equation (6) in a chimeric group will always be less than the value in equation (7) (*i.e.* $x_G < \theta$ for all chimeric groups). Therefore, we can evaluate the extent of the tragedy of the commons by considering how far groups are from the level of stalk production that maximises group success (eqn. 7).

Imperfect information

For strains to be able to make the optimal contribution to the stalk predicted in equation (6), they need perfect information of their relatedness to the group (r_i). In real biological systems, individuals are likely to use cues that provide information about relatedness, and therefore we do not expect them to have access to perfect information (*i.e.* they make errors in estimation). There are two levels of error that we account for in our analysis; the overall level of error or 'noise' in estimating relatedness, and the shape of the frequency-dependant error, where strains make smaller errors in estimating their relatedness to the group when they are at either very low or very high relatedness to the group.

To allow for error in estimating relatedness, we assume that cells within strains make unbiased normally distributed errors in their measurement of relatedness, with the mean of this distribution being their true relatedness to the group and the standard deviation corresponding to the degree of error in the measurement of relatedness (which reflects cells

over- or under-estimating their relatedness to the group). We model the standard error of this function by combining two processes, the inherent level of error in the measurement of relatedness (e) and the degree to which error is frequency dependent (f). For this we assign a level of error or 'noise', which ranges from 0 to 1, and then weight this value by a frequency dependent parameter (f), to give the standard deviation of the error function. Therefore, for any given level of inherent measurement error or noise (e), the realised level of error will depend on a strain's frequency in a group and the degree of frequency dependence of the error (f). The degree of frequency-dependence accounts for the fact that within-group heterogeneity (*i.e.* the variance in the genetic identity of the cells encountered within a group) depends on a strain's frequency, where the same level of inherent error (which reflects the biology of how relatedness is measured) is likely to realised errors of a larger magnitude in more heterogeneous groups. Therefore, we expect error to be highest when a strain is at a frequency of 0.5 and declines as they become either common or rare in groups. To allow for a range of shapes in this error function, we model it as a normal distribution with a mean of 0.5 and a standard deviation of f. In this way, the frequency dependent error parameter f captures the rate at which error drops as the frequency of a strain moves away from 0.5, with error dropping more rapidly with smaller values of f (see SI Appendix, Figure S2). To predict how a strain will behave given the level of error, we assume individual cells invest at a level given by the ESS strategy based on their measured relatedness to the group (rather than based on their true relatedness). As such, the level of investment by a strain at a given frequency reflects the averaging of the ESS strategy over the distribution of relatedness values measured by cells that are members of that strain.

Robustness to model assumptions

To provide a model that captures the most general scenario, we modelled benefits from the stalk as a linear function of investment. However, benefits could potentially be non-linear. Therefore, we evaluate the robustness of the main model predictions to non-linearity

of benefits using two general shapes of non-linear benefit functions: diminishing and accelerating returns. For each, we derive a new function for B_G (equation 2) and solve the ESS $\tilde{x}_i$ (equation 6). The non-linear equations for B_G are as follows: diminishing returns $B_G = 1 + b(1 - (1 - x_G)^{1.3}$ and accelerating returns $B_G = 1 + b{x_G}^{1.3}$ (see Figure 5A). Importantly, these different functions do not alter the qualitative pattern of investment by strains, but rather, they shift the expected level of investment above or below that expected from the linear function (see Figure 5). However, because these patterns are nearly identical to those expected from the linear benefits function with a different value of benefits and costs (shown by comparing the patterns in Figures 4&5), the model based on the assumption of linearity is used for fitting empirical data.

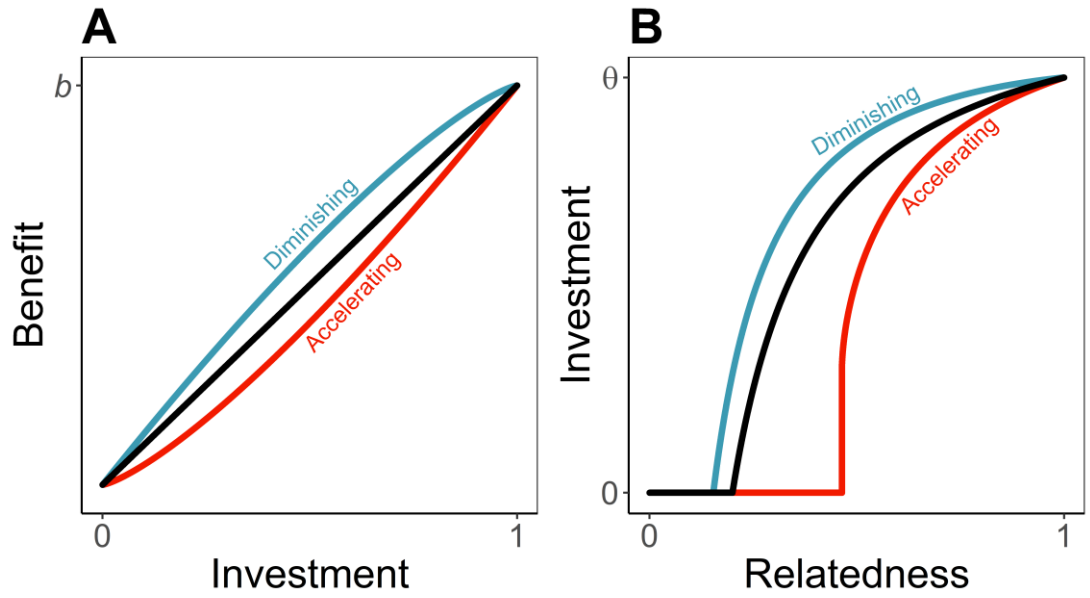


Figure 5. Linear and non-linear benefits of investment into public goods. **A**): Benefits of investment (black line) increase as investment increases. **B**) Optimal investment (black line) for one strain across the range of relatedness to the group. For both panels, the red and blue lines show data for diminishing and accelerating returns on investment respectively.

Empirical methods

We tested model predictions using a set of 24 naturally occurring strains of *D. discoideum* from North Carolina (NC), which have previously been used in many studies (Buttery *et al.*, 2009; Wolf *et al.*, 2015; Madgwick *et al.*, 2018): NC28.1, NC34.1, NC34.2, NC39.1, NC43.1, NC52.3, NC54.2, NC58.1, NC60.1, NC60.2, NC63.2, NC67.2, NC69.1, NC71.1, NC73.1, NC76.1, NC78.2, NC80.1, NC85.2, NC87.1, NC88.2, NC96.1, NC99.1, NC105.1.

Manipulation of group composition

In the first set of experiments we created groups containing three different strains. Because there is a huge array of possible frequency combinations that can be constructed from sets of three strains, we explored frequency space by varying frequencies along 'transects' through this space, where each strain was held constant at a frequency r_i of either 0.2 or 0.8 while the frequencies of two other strains in the group were varied across a set of ten frequency combinations (see Table 1). This yielded a total of 60 unique frequency combinations for a set of three strains (which represent 20 different frequency combinations, with each of three strains in a 'triplet' being treated as the focal strain in turn). These frequency combinations are indicated by the positions of the data points in Figure 1 across the three-strain frequency space. A total of three distinct triplets of strains were used, with each triplet replicated three times in each of the combinations, giving a total of $n = 540$ chimeric combinations. Each strain was also measured clonally three times in each replicate giving a total of $n = 540$ clonal measurements. In the second set of experiments we explored a wider range of relatedness values by increasing the number of strains in each group. For this we created groups of N strains (where N was 3, 4, 5, 6, 7, 8, 9, 10, 15, or 20) in which all strains were at a frequency of $1/N$ in the group, which means that the average relatedness in each group is $1/N$. Each of the ten conditions were replicated an average of 5.1 times, for a total of $n = 510$ chimeric combinations.

Table 1. Combinations of frequencies used in three-player experiments. Each column represents one of ten combinations for the two treatments, where one strain was held at a relatedness $r_1 = 0.2$ or 0.8.

Transect 1: $r_1 = 0.2$										
r_2 =	0.79	0.03	0.06	0.71	0.67	0.63	0.20	0.24	0.49	0.40
r_3 =	0.01	0.77	0.74	0.09	0.13	0.17	0.60	0.56	0.31	0.40
Transect 2: $r_1 = 0.8$										
r_2 =	0.01	0.18	0.17	0.04	0.05	0.14	0.13	0.12	0.09	0.10
r_3 =	0.19	0.02	0.03	0.16	0.15	0.06	0.07	0.08	0.11	0.10

Measurement of spore allocation

The protocols for quantifying spore allocation in *D. discoideum* are well-documented (*e.g.* see Kessin, 2001; Buttery *et al.*, 2009) and so described only briefly here. Strains were grown on *Klebsiella aerogenes* as a food source. After growth, amoebae were harvested and washed by centrifugation in KK2 buffer (16.1 mM KH_2PO_4 & 3.7 mM K_2HPO_4). Amoebae were then counted on a haemocytometer and resuspended in KK2 at a density of 10^8 cells per ml. Chimeric or clonal groups were created by adding cells from each strain at the relevant relative frequency in a 1.5ml Eppendorf and mixing thoroughly. 10^7 cells of each mix were then spread evenly on a 6cm petri dish containing 1.5% nutrient-free agar in KK2 and left to develop for 24 hours in an incubator at 22ºC. For collective investment measures (see below), all fruiting bodies were harvested in 5ml of spore buffer and counted on a haemocytometer. The total number of spores gave a measure of T_G (chimeric groups) or T_i (clonal groups) to be used for quantifying investment.

Measurement of fruiting body stability

To measure fruiting body stability we created groups of *N* strains following the same approach as described above in which all strains were at a frequency of $1/N$ in the group, with N = 3, 4, 5, 6, 7, 8, 9, 10, or 20). For fruiting body collapse we simply counted the total number of fruiting bodies on a plate, and the number of fruiting bodies that had collapsed after 24 hours. The percentage of total fruiting bodies that had collapsed was the measure of fruiting body collapse used. Each of the ten conditions were replicated an average of 4.4 times, for a total of $n = 440$ chimeric combinations.

Estimating contribution to public goods

The total level of collective investment by a group was estimated from the production of spores by the group, T_G, which reflects the inverse of allocation to the stalk (since cells not allocated to the stalk are necessarily allocated as spores). Because strains vary in their clonal level of investment in stalk, we normalized the behaviour of a strain in chimera to its behaviour

when clonal. For this, we measure spore production by a strain in clonal groups (T_i) and used this to calculate the expected spore production of a group as the weighted average of the clonal behaviour of each strain (where the clonal spore production by strains is weighted by their frequency in the group):

$$E_G = \sum_{i=1}^{N} r_i T_i \tag{8}$$

Because we expect strains to be investing at a level that corresponds to the optimum (θ), the value of E_G is expected to reflect the level of spore production when strains are investing at their optimal levels and hence provides an estimate of $1 - \theta$ in the model. A measure of the spore production was then calculated by comparing the measured spore production of the group, T_G, with the expectation if all strains were acting the same as they do when they develop grown clonally E_G:

$$S_G = \frac{T_G}{E_G} \tag{9}$$

Therefore, S_G will be greater than one if a chimeric group produces more spores than clones, which reflects a shift of cells away from stalk production into spore production. Because spore production is the inverse of allocation to stalk, this measure can be linked to the collective investment measure from the model (x_i) by simply taking the inverse of relative spore production.

$$I_G = \frac{E_G}{T_G} \quad (10)$$

This expression implies that, when $I_G = 1$ strains are showing the optimal level of investment (θ), and values less than one represent a reduction in stalk investment compared to the clonal expectation E_G.

Comparing empirical data to model predictions

To allow a direct comparison between model predictions and empirical data, we calculated collective investment from the model using the same method which is used to estimate the measure from empirical data. For this we need to express collective investment in terms of spore allocation scaled to the clonal expectation (equation 10). Therefore, the predicted collective investment from the model (x_G; eqn. 1) has an expected value (denoted X_G) when rescaled to match the empirical methods:

$$X_G = \frac{1-\theta}{\sum r_i(1-\tilde{x}_i)} \quad (11)$$

Statistical Analysis

Collective investment and fruiting body collapse across different number of strains (with all strains at equal relatedness) was modelled using mixed models fitted by maximum likelihood. In each model, group (which identified each unique combination of strains) was fitted as a random effect to control for variation in the collective behaviour of different strain combinations, and significance was assessed from an ANOVA of two models that differed only in the presence of absence of the 'number of strains' effect.

To test the robustness of the general patterns of decreased contribution to public goods as a function of relatedness, we test the effects of changing model parameters. One such key parameter is the 'strength of selection' S on the public good, which can be derived as the rate at which group fitness (equation 4) declines as a function of collective investment: $S = bc$. In order to vary the strength of selection while holding the optimal level of collective investment constant we simply rearranged equation (5) to solve either for *b* or *c*, and then solved the ESS (equation 9). We derived model predictions for a total of twelve different combinations of costs and benefits (with corresponding variation in the strength of selection S=2.98, 3.56, 4.27, 5.14, 6.25, 7.68, 9.59, 12.25, 16.17, 22.5, 34.32, 64).

In order to test the ability of the model to make quantitative predictions of collective investment (and therefore the tragedy of the commons) we searched for the set of parameters (*b, c, e,* and *f*) that provided the best fit between the given model's prediction of collective investment, and the empirical data for collective investment. When fitting models, we accounted for overall scale issues by allowing a small intercept (*I*) in the model, that is added to all empirical datapoints. The best-fit model has a small intercept (I =0.03) that improves the fit compared to a model with no intercept.

To find the best fit model, we firstly defined a large search-space of the four variables, *b, c, e,* and *f* ($2 \leq b \leq 10; 1 \leq c < 2; 0 \leq e \leq 1;\ 0 < f \leq 1$) with 20 values chosen for each parameter. For each unique combination of *b, c, e,* and *f,* we then derived the optimal strategy (equation 6) for each combination of *b, c, e,* and *f*. We then used our approach to modelling error (see 'Imperfect Information' section) to calculate the optimal strategy with error. Individual investment was converted to predicted collective investment (equation 11) corresponding to each of the unique social scenarios (relative relatedness of all players) for which we had empirical data. The fit between the prediction and the empirical data was assessed using a least-squares approach. To test the quality of the model fit to data, we used a paired t-test of each pair of empirical data and corresponding model prediction.

To calculate confidence intervals around the best-fit model we use a resampling approach. Briefly, we took samples of each datapoint from distributions corresponding to the empirical data, and calculated the best-fit parameters using the same approach as above. We then repeated this approach for $n = 100$ iterations. To calculate confidence intervals for each parameter, we used the range between the 5th and 95th percentiles of the deviations between the parameter value of the overall best-fit model and the parameter value of each iteration. For plotting, we calculate a single confidence interval around the overall best fit as the combination of upper and lower confidence intervals of all three parameters that gives the greatest deviation in predicted investment from the overall best-fit model. As such, our confidence interval is a conservative estimate of confidence in the true values of the parameters.

To assess the utility of adding another variable, error in relatedness estimation, to the model fit, we first used a broad search-space of parameters b and c and a least-squares approach to fit the perfect information model with the b and c values that best matched the data. Next, we added another parameter, e, to make an imperfect information model, where e is the standard deviation of the error in measuring relatedness (see 'Imperfect information' section). To test for the significant of difference in fit between perfect and imperfect information models, we used an F-test.

Acknowledgements

This work was supported by a Natural Environment Research Council (NERC) GW4+ PhD Studentship to LJB, a Biological Sciences Research Council (BBSRC) SWBio PhD Studentship to PGM, and a grant from the BBSRC (BB/M01035X/1) to CRLT and JBW.

References

Ågren, J.A., Davies, N.G. & Foster, K.R. 2019. Enforcement is central to the evolution of cooperation. *Nat. Ecol. Evol.* **3**: 1018-1029.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E.A., et al. 2009. Polymorphic members of the lag gene family mediate kin discrimination in Dictyostelium. *Curr. Biol.* **19**: 567–72.

Bourke, A.F.G. 2011. *Principles of Social Evolution*. Oxford University Press, Oxford.

Boyd, R., Gintis, H. & Bowles, S. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science (80-.).* **328**: 617–620.

Breed, M.D., Welch, C.K. & Cruz, R. 1994. Kin discrimination within honey bee (Apis mellifera) colonies: An analysis of the evidence. *Behav. Processes* **33**: 25–39.

Bshary, R. & Grutter, A.S. 2005. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biol. Lett.* **1**: 396–399.

Buttery, N.J., Rozen, D.E., Wolf, J.B. & Thompson, C.R.L. 2009. Quantification of Social Behavior in D. discoideum Reveals Complex Fixed and Facultative Strategies. *Curr. Biol.* **19**: 1373–1377.

Charnov, E.L. 1977. An elementary treatment of the genetical theory of kin-selection. *J. Theor. Biol.* **66**: 541–550.

Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature* **373**: 209–216.

Dickinson, J.L. 2004. A test of the importance of direct and indirect fitness benefits for helping decisions in western bluebirds. *Behav. Ecol.* **15**: 233–238.

Dionisio, F. & Gordo, I. 2007. Controlling excludability in the evolution of cooperation. *Evol. Ecol. Res.* **9**: 365–373.

Dionisio, F. & Gordo, I. 2006. The Tragedy of the Commons, the Public Good Dilemma and the Maning of Rivalry and Excludability in Evolutionary Biology. *Evol. Ecol. Res.* **8**: 321–332.

Duncan, C., Gaynor, D., Clutton-Brock, T. & Dyble, M. 2019. The Evolution of Indiscriminate Altruism in a Cooperatively Breeding Mammal. *Am. Nat.* **193**: 841–851.

Ferrari, M. & König, B. 2017. No evidence for punishment in communally nursing female house mice (Mus musculus domesticus). *PLoS One* **12**: 1–16.

Ferrari, M., Lindholm, A.K. & König, B. 2015. The risk of exploitation during communal nursing in house mice, Mus musculus domesticus. *Anim. Behav.* **110**: 133–143.

Foster, K.R., Shaulsky, G., Strassmann, J.E., Queller, D.C. & Thompson, C.R.L. 2004. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**: 693–696.

Foster, K.R., Wenseleers, T. & Ratnieks, F.L.W. 2006. Kin selection is the key to altruism. *Trends Ecol. Evol.* **21**: 57–60.

Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520–522.

Frank, S.A. 1996. Policing and group cohesion when resources vary. *Anim. Behav.* **52**: 1163–1169.

Frank, S.A. 2009. Evolutionary Foundations of Cooperation and Group Cohesion. In: *Games, Groups, and the Global Good* (S. Levin, ed), pp. 3–40. Springer Physica-Verlag Berlin Heidelberg.

Gardner, A. & West, S.A. 2004. Cooperation and Punishment, Especially in Humans. *Am. Nat.* **164**: 753–764.

Gardner, A. & West, S.A. 2014. Inclusive fitness: 50 years on. *Philos. Trans. R. Soc. B Biol. Sci.* **369**: 20130356.

Gordon, H.S. 1954. The Economic Theory of a Common-Property Resource: The Fishery. *J. Polit. Econ.* **62**: 124–142.

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J.A., Wolf, J.B. & Thompson, C.R.L. 2017. A polychromatic "greenbeard" locus determines patterns of cooperation in a social amoeba. *Nat. Commun.* **8**: 1–9.

Hamilton, W.D. 2001. *Narrow Roads of Gene Land: Volume 2*. Oxford University Press, Oxford.

Hamilton, W.D. 1964. The Genetical Evolution of Social Behaviour. II. *J. Theor. Biol.* **7**: 17–52.

Hardin, G. 1968. The tragedy of the commons. *Science (80-.).* **162**: 1243–1248.

Keller, L. 1997. Indiscriminate altruism: Unduly nice parents and siblings. *Trends Ecol. Evol.* **12**: 99–103.

Kessin, R.H. 2001. *Dictyostelium: Evolution, Cell Biology, and the Development of Multicellularity*. Cambridge University Press.

Komdeur, J., Richardson, D.S. & Burke, T. 2004. Experimental evidence that kin discrimination in the Seychelles warbler is based on association and not on genetic relatedness. *Proc. R. Soc. B Biol. Sci.* **271**: 963–969.

Konig, B. 1994. Fitness effects of communal rearing in house mice: the role of relatedness versus familiarity. *Anim. Behav.* **48**: 1449–1457.

Madgwick, P.G., Stewart, B., Belcher, L.J., Thompson, C.R.L. & Wolf, J.B. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. *Proc. Natl. Acad. Sci.* **115**: E4823–E4832.

Manhes, P. & Velicer, G.J. 2011. Experimental evolution of selfish policing in social bacteria. *Proc. Natl. Acad. Sci.* **108**: 8357–8362.

Morris, H.R., Taylor, G.W., Masento, M.S., Jermyn, K.A. & Kay, R.R. 1987. Chemical structure of the morphogen differentiation inducing factor from Dictyostelium discoideum. *Nature* **328**: 811–814.

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

Nonacs, P. 2011. Kinship, greenbeards, and runaway social selection in the evolution of social insect cooperation. *Proc. Natl. Acad. Sci.* **108**: 10808–10815.

Olson, M. 1965. *The Logic of Collective Action*. Harvard University Press.

Ostrom, E. 1990. *Governing the Commons: the evolution of institutions for collective action*. Cambridge University Press.

Ostrowski, E.A., Katoh, M., Shaulsky, G., Queller, D.C. & Strassmann, J.E. 2008. Kin Discrimination Increases with Genetic Distance in a Social Amoeba. *PLoS Biol.* **6**: e287.

Parkinson, K., Buttery, N.J., Wolf, J.B. & Thompson, C.R.L. 2011. A Simple Mechanism for Complex Social Behavior. *PLoS Biol.* **9**: e1001039.

Pellmyr, O. & Huth, C.J. 1994. Evolutionary stability of mutualism between yuccas and yucaa moths. *Nature* **372**: 257–260.

Rankin, D.J., Bargum, K. & Kokko, H. 2007. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**: 643–651.

Ratnieks, F.L.W. & Reeve, H.K. 1992. Conflict in single-queen hymenopteran societies: the structure of conflict and processes that reduce conflict in advanced eusocial species. *J. Theor. Biol.* **158**: 33–65.

Sasaki, T. & Uchida, S. 2014. Rewards and the evolution of cooperation in public good games. *Biol. Lett.* **10**: 20130903.

Smith, J., Queller, D.C. & Strassmann, J.E. 2014. Fruiting bodies of the social amoeba Dictyostelium discoideum increase spore transport by Drosophila. *BMC Evol. Biol.* **14**: 105.

Strassmann, J.E. & Queller, D.C. 2014. Privatization and property in biology. *Anim. Behav.* **92**: 305–311.

Strassmann, J.E., Zhu, Y. & Queller, D.C. 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum. *Nature* **408**: 965–967.

Taylor, P.D., Wild, G. & Gardner, A. 2007. Direct fitness or inclusive fitness: How shall we model kin selection? *J. Evol. Biol.* **20**: 301–309.

Travisano, M. & Velicer, G.J. 2004. Strategies of microbial cheater control. *Trends Microbiol.* **12**: 72–78.

Trivers, R.L. 1971. The Evolution of Reciprocal Altruism. *Q. Rev. Biol.* **46**: 35–57.

Van Dyken, J.D., Linksvayer, T.A. & Wade, M.J. 2011. Kin Selection–Mutation Balance: A Model for the Origin, Maintenance, and Consequences of Social Cheating. *Am. Nat.* **177**: 288–300.

Velicer, G.J. 2003. Social strife in the microbial world. *Trends Microbiol.* **11**: 330–337.

Wang, Y. & Shaulsky, G. 2015. TgrC1 Has Distinct Functions in Dictyostelium Development and Allorecognition. *PLoS One* **10**: e0124270.

Wechsler, T., Kümmerli, R. & Dobay, A. 2019. Understanding policing as a mechanism of cheater control in cooperating bacteria. *J. Evol. Biol.* **32**: 412–424.

Wenseleers, T., Helanterä, H., Hart, A. & Ratnieks, F.L.W. 2004. Worker reproduction and policing in insect societies: An ESS analysis. *J. Evol. Biol.* **17**: 1035–1047.

Wenseleers, T. & Ratnieks, F.L.W. 2004. Tragedy of the commons in Melipona bees. *Procedings R. Soc. B Biol. Sci.* **271**: S310–S312.

West, S.A., Kiers, T.E., Pen, I. & Denison, R.F. 2002. Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15**: 830–837.

West, S.A.S.A., Griffin, A.S. & Gardner, A. 2007. Evolutionary Explanations for Cooperation. *Curr. Biol.* **17**: R661-R672.

Wolf, J.B., Howie, J.A., Parkinson, K., Gruenheit, N., Melo, D., Rozen, D., et al. 2015. Fitness Trade-offs Result in the Illusion of Social Success. *Curr. Biol.* **25**: 1086–1090.

Comment 2: Constraints on conditional cooperation

The preceding paper provided an analysis of how non-adaptive constraints can lead to beneficial outcomes for chimeric mixes of *Dictyostelium discdoideum*. Here, I use these findings to speculate on the possible mechanism through which strains are able to engage in conditional cooperation.

A classic expectation of mechanisms of social discrimination is that these mechanisms would evolve such that social partners can be recognised with lower error (Hamilton, 1964b). Yet, a 'veil of ignorance' (Rawls, 1971) that inhibits recognition mechanisms could be selected under special conditions (Ridley, 2000), such as the meiotic scrambling of information about which genes are passed on to offspring under sexual recombination (Haig & Grafen, 1991). For a veil of ignorance to be favoured, an individual must benefit from reducing their own level of information, which would only occur if, concurrently, competitors also had a reduced level of information. In this way, an unconditional strategy of cooperation could only selectively replace a conditional strategy of cooperation if the unconditional strategist can remove the advantages that information provides to a conditional strategist. Although conditional cooperation may be less cooperative on average, the ESS level of cooperation of an unconditional strategist may not be very high (depending on the probability distribution of frequencies that a player experiences). Consequently, the dilemma within the tragedy of the commons cannot be escaped by unconditional strategy – a point that is often missed (*e.g.* Queller & Strassmann, 2013). Therefore, whilst it is true that more perfect social discrimination is likely to facilitate a more effective conditional strategy, better information does not always benefit individuals because it can exacerbate the tragedy of the commons.

Paper 3 showed that biological constraints that limit the ability of an individual to modulate their conditional strategy can increase the level of cooperation between individuals. These constraints are non-adaptive, for it remains true that if an individual could overcome them then it would reap an advantage – though this advantage would be short-lived (until the

better-informed genotype has gone to fixation) and ultimately detrimental to group fitness (in exacerbating the tragedy of the commons). In *D. discoideum*, strains play their strategy of conditional cooperation under constraints: informational constraints wherein a strain is not able to detect their true frequency within the aggregate, developmental constraints wherein a strain is not able to freely change how their cells differentiate, and selective constraints wherein a strain's strategy is not selected to optimise their fitness at a particular frequency. The selective constraint is potentially interrelated with the other forms of constraint, but also has an independent quality even if selection is unable to improve a strain's ability to detect their frequency or change how cells differentiate. All these different sources of constraint are likely to have some role, but the evidence presented in Paper 2 suggests that informational constraints are probably the most important because only high error in the ability to detect focal strain frequency produces the characteristic pattern of investment behaviour in the data.

Although there are numerous biochemical mechanisms through which a strain could detect its frequency within an aggregate, the evidence in Paper 2 suggests that strains do not respond to the number or frequency of other players. In other words, strains appear to use self-referential information only. Consequently, there are two basic options for signalling this information: a secreted molecule and a cellular receptor or a cell-surface signal-receptor. A candidate secreted signal might be DIF-1, which is known to induce stalk cell differentiation in *D. discoideum* (Thompson & Kay, 2000). In principle, strains could direct their own stalk cell differentiation whilst ignoring the DIF signal in a competing strain (Parkinson *et al.*, 2011). However, there is no known polymorphism in any of the DIF molecules in *D. discoideum*. Perhaps there are other secreted molecules that are yet unknown, but secreted molecules might readily provide some selective pressure for interfering with other strains by producing both a self-signal and another nonself-signal to disrupt the competing strain, and there is no evidence of overproduction of stalk (compared to the clonal level) by any strain combination. Alternatively, a candidate cell-surface signal might be TgrB1/C1, which are a linked pair of cell binding proteins that each bind the other. Previous analysis has shown that

Tgr protein binding strength predicts segregation behaviour (Gruenheit *et al.*, 2017), which Paper 1 showed to be frequency-dependent. The Tgr proteins are likely to cause segregation through the direct action of cell binding, as groups of cells are more likely to remain in the same aggregate if they more strongly bind one another. Although the process of cellular differentiation is likely to be more complicated than cell binding (*i.e.* involving many other genes), similar proteins could be involved in a signalling cascade that leads to downstream effects. The *tgr* gene family is very large, with numerous duplicated genes and pseudogenes (but unpublished data shows that *tgrb1/c1* does not explain frequency-dependent investment behaviour). Nonetheless, the *tgr* loci are amongst the most polymorphic loci in *D. discoideum*'s genome, and are the only signal-receptors amongst the genes with high enough polymorphism to explain frequency-dependent investment behaviour in the test-set of naturally co-occurring strains (Gruenheit *et al.*, 2017; de Oliveira *et al.*, 2019). Therefore, I would suggest that it is most likely that a self-binding protein is responsible for how cells detect their frequency within an aggregate.

Whether the signalling relies on a secreted or cell-surface mechanism, a major evolutionary problem for understanding how conditional cooperation is possible in *D. discoideum* is how selection favours strains with different self-signals (which defines whether a cell belongs to one strain or other). It is the general expectation that self-signals have a common-type advantage; a common self-signal should go to fixation because its frequency means that is more likely to interact with other variants that share the same self-signal, and therefore it is simply more likely to engage in cooperation. This problem – known as Crozier's paradox – would suggest that all *D. discoideum* (or at least all naturally co-occurring types) would effectively belong to one strain because they all bear the same self-signal.

Paper 2 has demonstrated how many of the predictions of the Collective Investment game are inaccurate because of constraints in the social system of *D. discoideum*. Comment 2 has speculated on what the deviations from predictions imply about the mechanism of conditional cooperation in *D. discoideum*. Next, I move onto considering the Collective

Investment game as a generalised framework for understanding conditional cooperation, with a view to finding a solution to Crozier's paradox in this specific context.

Paper 3

This declaration concerns the article entitled:			
Evolution of strategic cooperation			
Publication status (tick one) **Draft manuscript** ☐ **Submitted** ☐ **In review** ✓ **Accepted** ☐ **Published** ☐			
Publication details (reference)	**Madgwick, P. G.**, and J. B. Wolf. n.d.. Evolution of strategic cooperation.		
Copyright status (tick the appropriate statement) I hold the copyright for this material ✓ Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Formulation of ideas: 80%: PGM significantly contributed to the concept, literature review and narrative opinion. Design of methodology: N/A Experimental work: N/A Presentation of data in journal format: 70%: PGM predominantly executed the drafting, revising and editing of the manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	PGMadgwick	**Date**	07/11/2019

Data access statement: email pgm29@bath.ac.uk

Evolution of strategic cooperation

Authors: Philip G. Madgwick[1*], and Jason B. Wolf[1]

Affiliations: [1] Milner Centre for Evolution and Department of Biology & Biochemistry, University of Bath, Bath, BA2 7AY, UK

[*] Correspondence to: pgm29@bath.ac.uk

Abstract

Group-beneficial behaviours have presented a long-standing challenge for evolutionary theory because, whilst their benefits are available to all group members, their costs are borne by individuals. Consequently, an individual could benefit from 'cheating' their group-mates by not paying the costs while still reaping the benefits. There have been many proposed evolutionary mechanisms that could favour cooperation (and disfavour cheating) in particular circumstances. However, if cooperation is still favoured in some circumstances then we might expect evolution to favour strategic cooperation, where the level of contribution toward group-beneficial behaviour is varied in response to the social context. To uncover how and why individuals should contribute toward group-beneficial behaviour across social contexts, we model strategic cooperation as an evolutionary game where players can quantitatively adjust the amount they contribute toward group-beneficial behaviour. We find that the evolutionarily stable strategy (ESS) predicts, unsurprisingly, that players should contribute in relation to their relatedness to the group. However, we surprisingly find that players often contribute to cooperation in such a way that their fitness is inverse to their relatedness to the group such that those that contribute to cooperation end up with the same return from group-beneficial behaviour, essentially removing any potential advantage of higher relatedness. These results bring to light a paradox of group-beneficial cooperation: groups do best when they contain highly related individuals, but those with the highest relatedness to the group will often have

the lowest fitness within the group.

Keywords: cooperation**,** social behaviour, public goods, kin selection, group selection

Introduction

Group-beneficial social behaviour is near-universal across the diversity of life, ranging from communal care in vertebrates (Clutton-Brock, 2009) through self-sacrificial defensive behaviours in eusocial insects (Wilson & Hölldobler, 1990) to secreted metabolic molecules in micro-organisms (West *et al.,* 2006). Despite covering wildly different taxa and ecologies, these types of social behaviours all share a common paradox: the benefits are available to everyone in the group, and yet their costs are borne by the particular individuals that perform them (Wilson, 1977; Frank, 1995; Rankin *et al.*, 2007; Johnstone & Rodrigues, 2016). Why should an individual pay the cost of group-beneficial behaviour rather than 'cheating' their group-mates by not paying their fair share of the costs? In general, the evolutionary problem of cheating is expected to lead natural selection to disfavour group-beneficial behaviours (Frank, 2003; Foster, 2004; Travisano & Velicer, 2004; West *et al.*, 2007a; Bourke, 2011), but there are numerous mechanisms that tip the balance of selection in their favour within specific contexts, such as high relatedness between social partners (Hamilton, 1964a), strong reciprocal interactions (Trivers, 1971), and enforcement mechanisms like punishment (Clutton-Brock & Parker, 1995) or policing (Frank, 1995).

Logically, if cheating is favoured in some circumstances and not in others, then we might expect selection to favour 'strategic' cooperation, where individuals modulate the amount they contribute toward group-beneficial behaviour in different situations. Although learning could potentially play a role in strategic cooperation, here we are focused on genetic strategies that evolve to optimally adjust the level of cooperation by an individual to suit its social context. Such strategic cooperation could appear as facultative cooperation, where

individuals contribute toward group-beneficial behaviour when it is profitable (in terms of fitness), whilst withholding contributions when it is not. However, strategic cooperation need not appear as an all or none pattern of behaviour. Indeed, an individual would often do better by strategically adjusting their contribution to cooperation in response to the relative costs and benefits. In this way, strategic cooperation is not a mechanism that prevents or avoids cheating but, rather, an adaptation that allows individuals to maximise their fitness within any given social context (whether this means making contributions toward cooperation or free-riding on the contributions of others).

Strategic cooperation is most likely to evolve when social contexts are highly variable, and hence natural selection can favour different levels of cooperation in different situations. For example, pairs of female house mice have been shown to adjust their contribution to communal care in response to unequal litter sizes, increasing the time spent caring for the group's offspring when their own offspring make up a larger fraction of the group (Ferrari *et al.*, 2016). Similarly, soldier-producing aphids increase their contributions toward gall defence behaviour when in highly related groups, which comes at a cost to their personal reproduction (Abbot *et al.*, 2001). Likewise, strains of a social amoebae modulate the fraction of cells that differentiate into non-reproductive stalk cells (that help the remaining spores cells to disperse) dependent upon their frequency within the group (Madgwick *et al.*, 2018). These examples, from cooperatively breeding vertebrates, eusocial insects and social micro-organisms, indicate that strategic cooperation can occur across diverse taxa and that it need not require complex cognitive skills. Consequently, strategic cooperation may help explain why group-beneficial behaviours persist throughout nature, despite the potential advantages of cheating. Strategic cooperation may also explain the 'missing cheaters' phenomenon – why obligate cheater types are so surprisingly rare in nature (Gilbert *et al.*, 2007). Instead, if individuals are able to show flexible social strategies where they strategically modulate their level of cooperation to best exploit their social environment, we would expect any individual to appear to be a contributor or a free-rider (*i.e.* appear to be a cheater) in different situations.

To uncover how individuals should change their contributions toward group-beneficial behaviour in different social contexts, we develop a general theoretical framework to model strategic cooperation as a flexible social strategy where individuals can quantitatively adjust their contributions toward group-beneficial 'public goods'. We identify the evolutionarily stable strategy (ESS), which represents a quantitative pattern of how much an individual should contribute towards cooperation across social contexts to maximise their fitness. To elucidate the factors that shape the ESS, we examine the evolutionary logic of strategic cooperation within different conceptual frameworks (kin and group selection).

Models and Results

We model cooperation through a group-beneficial behaviour that generates a 'public good' that benefits all group members (Hamburger, 1973; Fox & Guyer, 1978; Dionisio & Gordo, 2006; Frank, 2010). Hence, our analysis does not include cooperative behaviours that are limited to one-way transactions, such as where an actor either helps or harms a recipient. Across different scenarios, the prevailing feature of public goods is that they are costly to produce, leading to a trade-off between the production of public goods and other fitness related traits (Haldane, 1932; Wright, 1945; Maynard Smith, 1964; Wilson, 1975a; Frank, 1995). Cooperation through production of costly public goods is widespread, but the form of cooperation can vary widely. Public goods could refer to a physical product, such as secreted molecules in microbes that come at an energetic cost to the cells that produce them (which presumably reduces the growth rate of those cells), while being available to all cells in the local environment (West *et al.*, 2006). For example, iron-scavenging siderophores in the bacterium *Pseudomonas aeruginosa* (Griffin *et al.*, 2004) and extracellular enzymes within biofilms in the virus *Vibrio cholerae* (Drescher *et al.*, 2014) are energetically costly to produce and increase the fitness of all cells in the local area. Similarly, individuals can cooperate in the production of physical structures that act as public goods, such as the fruiting body stalk that

facilitates dispersal in the social amoeba *Dictyostelium discoideum* (Madgwick *et al.*, 2018) and social bacteria *Myxococcus xanthus* (Velicer *et al.*, 2000), which comes at a clear cost to the cells that are sacrificed for its production. Public goods could also arise from behavioural services (Foster, 2004), such as vigilance in meerkats (Santema & Clutton-Brock, 2013) and monkeys (Gaynor & Cords, 2012), which is costly because it competes with time that an animal could use for other purposes and may also expose the individual to direct danger, while benefiting all individuals in the local area who can respond to an alarm call (signalling an approaching predator). Likewise, cooperatively breeding animals may often generate public goods through indiscriminate communal care (Riehl, 2013; Ferrari *et al.*, 2015) that benefit all the young in a communal nest, while coming at a cost to an individual's own fecundity. Thus, given that there is a broad range of scenarios where individuals cooperate through public goods in nature, we do not construct our model to match any specific biological scenario. Instead, we implement a generalised framework (building on ref 18) that considers the problem from the perspective of competing genetic variants that are maximising their transmission, which can be adapted to fit a large range of different systems.

We build a game-theoretic model where the 'players' are genetic variants (*i.e.* coreplicons; Cosmides & Tooby, 1981). We consider groups of N players, where each (i^{th}) player is present at a particular frequency p_i within the group (such that $\sum_{i=1}^{N} p_i = 1$). We assume that a player's frequency within a group is independent of their strategy and that variation in the frequency of a player across groups is caused by some random process that allows players to potentially experience a range of group compositions (and so we examine the consequences of variation in group composition, not the underlying causes of the variation). In the absence of any contribution to public goods, all players within a group have equal fitness that is arbitrarily assigned a value of 1, which can be viewed as the budget from which they can make a contribution toward public goods. In this way, we assume the cost of producing public goods is in units of fitness, and hence making a contribution comes at the expense of a player's potential fitness through other traits. Players contribute a proportion (x_i)

of this potential fitness into public goods ($0 \leq x_i \leq 1$), which results in a reduction in fitness of c per unit contributed, making the total cost cx_i. Because contributions are measured in units of fitness, logically c would equal 1. However, by assigning the total cost as cx_i we are able to capture scenarios where fitness declines at a rate that is more or less than the linear expectation across the relevant range of contributions (*i.e.* between zero contribution and the optimal level of contribution that maximizes fitness, which should typically capture a range far below a contribution of 1). After contributing to public goods, a player has a residual fitness of $1 - cx_i$ before accounting for any fitness benefits from the public goods. The benefit from public goods depends on the total collective contribution made by all members of a group ($x_G = \sum_{i=1}^{N} p_i x_i$) since the public goods benefit everyone, regardless of which player 'paid' to produce them. Each unit of collective contribution gives a benefit of b in units of fitness, which makes the total impact of collective contribution to public goods on a player's fitness $1 + bx_G$. Although it is possible that some systems might show a non-linear relationship, such as those with diminishing or accelerating benefits from public goods, the general patterns of results would hold under these other relationships insofar as they are monotonically increasing functions (and, moreover, even non-linear relationships may be quasi-linear in the evolutionarily relevant range of contributions). Thus, a player's realised fitness (after accounting for the costs and benefits of public goods) is determined by the multiplicative effects of the costs from contributing to $(1 - cx_i)$ and the benefits from the availability of public goods $(1 + bx_G)$:

$$\omega_i = (1 - cx_i)(1 + bx_G) \quad (1)$$

Hence, a player's fitness is essentially the realised benefit, which depends on both the fitness cost paid for their contribution to and benefit arising from public goods (since the two are multiplicative fitness components); it is through the residual potential fitness after contributing

that players gain the fitness benefit from the public goods and so contributing to public goods reduces a player's potential benefit from the public goods.

Table 1. Notation for the model.

Parameter	Definition
b	Benefit (per unit of collective investment) to the group from public goods
c	Cost (per unit of investment) to a player from public goods
N	Number of players in the group
ω_i	Fitness of the i^{th} player
p_i	Frequency of the i^{th} player
x_i	Investment strategy of the i^{th} player
x_G	Collective investment of all players in the group
$\hat{x}_i$	Evolutionarily stable strategy (ESS) with respect to the i^{th} player
n	Number of players in the group that contribute to public goods
$\bar{p}$	Average frequency of all contributing players in the group
$\hat{x}_G$	Coevolutionarily stable strategy (coESS) of collective investment by a group of players each pursuing the ESS

Although a player's fitness depends on the total collective contributions to public goods made by the group, to understand why a player personally contributes we need to consider their marginal impact on their own fitness owing to their own contribution to collective investment: $1 + bp_i x_i$. Given that contributing to public goods comes at a cost with respect to potential fitness $1 - cx_i$, any contribution by a player must ultimately increase their fitness above their baseline fitness: $(1 + bp_i x_i) > 1/(1 - cx_i)$. This perspective emphasizes the dilemma of cooperation through public goods: the cost is paid directly, but benefits are diluted across the group, and therefore a genetic variant only sees a return on its contribution to public goods in relation to its frequency in the group.

The biologically-relevant solution for players' investment strategies is the evolutionarily stable strategy (ESS), which describes the best strategy that a player can adopt in an equilibrium population of strategies that have evolved by natural selection (Maynard Smith, 1974). An additional level of complexity for the ESS of a continuously-variable social behaviour is that the ESS must describe a player's strategy across the full range of social contexts – and, critically, with respect to whatever information causes that player to modulate their strategy. Here we focus our attention on the most unconstrained case, where the ESS describes what proportion of a player's fitness budget they sacrifice to produce public goods across groups with any number of players and any distribution of frequencies of players within the group. Thus, we assume that players have access to perfect information about their social context. By considering players to be genetic variants, this assumption implies that they are able to measure their frequency within the group (*e.g.* using a signal) and can then modify their contribution toward public goods in response (which is a necessary requirement of any frequency-dependent strategy). We further assume that each player's response to a given social context is independent of its response to other social contexts, which simply means that the pattern of cooperation across social contexts is genetically unconstrained and can evolve to maximize a genetic variant's fitness (Maynard Smith, 1976; Grafen, 1984; Parker & Maynard Smith, 1990; Kirkpatrick & Gomulkiewicz, 1992).

Using optimality assumptions (Maynard Smith, 1976; Grafen, 1984; Parker & Maynard Smith, 1990; Kirkpatrick & Gomulkiewicz, 1992), we can solve the ESS (which we denote $\hat{x}_i$) by finding the pattern of contribution to public goods that maximizes a player's fitness (ω_i) given their frequency in a group. We refer to this pattern as their 'investment strategy' to reflect the fact that the level of contribution to public goods is that which maximises the benefit against the costs. Given that a player's investment strategy is defined as the proportion of their potential fitness that they devote towards public goods, the ESS can be given with respect to some equilibrium quantity of collective investment by the group ($\hat{x}_G$) by setting the derivative of Equation (1), $d\omega_i/dx_i$), equal to zero and rearranging (see also SI Appendix, Section 1):

$$\hat{x}_i = \frac{p_i b - c - bc\hat{x}_G}{p_i bc} \tag{2}$$

In broad terms, the ESS shows that a player contributes a larger proportion of their fitness budget toward public goods when they are at a higher frequency in a group (p_i), but their exact level of investment can depend on the behaviour of others (which is reflected in the dependence of individual investment on group investment). Because the solution to equation (2) depends on the level of investment by the group ($\hat{x}_G$; *i.e.* including the focal player) it does not represent a closed solution (since the investment made by each group member relies on the level made by all other group members, and hence the values are all interdependent). Therefore, to solve a player's ESS level of investment towards public goods, the expression in equation (2) must be resolved simultaneously for all players in a group.

Across all possible values of the benefit and cost terms (b, c), we know of no method that can analytically solve the ESS, so we implement numerical analysis (Figure 1). However, it is possible to derive a more exact analytical description of the ESS by further resolving the quantity of collective investment by the group ($\hat{x}_G$) when each player contributes toward public goods at the level of the ESS for their frequency, which we refer to as the

coevolutionarily stable strategy (coESS). The coESS is the sum of the contributions made by all group members toward public goods at equilibrium across the full range of possible frequencies of players within groups. Within any particular group, not all players necessarily contribute toward public goods at the coESS because, under many conditions, one or more players pursuing the ESS do not contribute given their frequency. Consequently, at the coESS the total investment depends on a subset of investing players; for any particular group, we can therefore define that there is a set of n players who invest, which we refer to as the 'contributors', and $N - n$ players that do not, which we refer to as the 'free-riders'. We can express the solution in terms of the mean of the frequencies of all contributors ($\bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i$), which also gives the proportion of the group that contributes $\bar{p}n$ (making the proportion of free-riders in a group $1 - \bar{p}n$). The total investment for a group is, therefore, the sum of the investment by contributors ($\hat{x}_G = \sum_{i=1}^{n} p_i \hat{x}_i$), which we can describe with greater specificity using these new terms:

$$\hat{x}_G = \frac{n(\bar{p}b - c)}{(n+1)bc} \tag{3}$$

In broad terms, the coESS shows that a group of players invest more into public goods when there are fewer contributors at a higher average frequency. Because the coESS reflects the ESS behaviour for a group of contributors, we can substitute the coESS into the ESS (eqn. 2) to arrive at a more exact analytical solution for the ESS level of investment by a contributor:

$$\hat{x}_i = \frac{p_i b - c - n(\bar{p} - p_i)b}{p_i(n+1)bc} \tag{4}$$

This solution to the ESS is only fully resolved for a particular group once players are classified as being one of the n contributors or $N - n$ free-riders. A player can be categorised as a contributor or free-rider based on their frequency in a group by finding the minimum

frequency that separates contributors from free-riders for a given group (within a numerical analysis; Table 2).

A
B
C
θ
0
p*
1
Level of Investment (x_i)
Player Frequency in the Group (p_i)

1

Figure 1. The ESS patterns of individual investment into public goods across frequencies (p_i) under different cost-benefit scenarios. In each panel the red line indicates the ESS for the focal player in groups composed of two players (*i.e.*, $\hat{x}_i$ when $N = 2$) and the blue line the ESS for the focal player in groups that contain a large number of non-focal players, each at low frequency. These lines overlap (indicated by the alternating red and blue dashed lines) when the focal players is at a frequency that is above the threshold p^* (indicated by light-grey shading), which is the frequency at which the focal player is the only contributor toward public goods. In regions where these two lines do not overlap, they set upper and lower limits to the expected level of investment (since players reduce investment in response to investment by others, groups with a large number of other players will set an upper limit on investment while groups with two players sets a lower limit on investment; see eqn. 6). The purple shaded area, therefore, indicates the range of possible patterns of investment by the focal player, with the exact value depending on the distribution of frequencies across non-focal players within the group. The solid black horizontal line indicates the level of investment that maximises group fitness (θ; eqn. 9). The three panels show patterns corresponding to different benefit-to-cost ratios (where costs were held constant at $c = 1$ and benefits b were varied), which were chosen to capture three fundamental scenarios: **A)** benefits relative to costs are low ($b = 3/2$), such that players only contribute when they have a high frequency and consequently there is only ever a single contributor, **B)** benefits relative to costs are high ($b = 3$) and consequently there is potentially a small overlap between conditions where the focal player contributes and non-focal players might also contribute, and **C)** benefits relative to costs are very high ($b = 9$) and consequently there is a very large region where non-focal players may be motivated to contribute.

Table 2. The categorization of players as free-riders or contributors. All players within a group can be classed as a contributor or free-rider depending on their frequency. The frequency threshold that separates the two classes depends on the costs (c) and benefits (b) of investment into public goods. Once classified, players can be assigned an investment level ($\hat{x}_i$) and expected fitness (ω_i).

Investment class	**Frequency limit**	**Investment level**	**Fitness**
Free-rider	$p_i < \frac{\bar{p}nb + c}{b(n+1)}$	$\hat{x}_i = 0$	$\omega_i = \frac{(\bar{p}nb + c)}{c(n+1)}$
Contributor	$p_i \geq \frac{\bar{p}nb + c}{b(n+1)}$	$\hat{x}_i = \frac{p_i b - c - n(\bar{p} - p_i)b}{p_i bc(n+1)}$	$\omega_i = \frac{(\bar{p}nb + c)^2}{p_i bc(n+1)^2}$

There is a fascinating and unexpected property of the ESS: each of the n contributors has equal transmission to the next generation, despite having different starting frequencies. Transmission refers to the number of copies of that particular genetic variant that are passed to the next generation, which is calculated as their frequency-weighted fitness ($p_i\omega_i$). Thus, in real terms, equal transmission would mean that genetic variants produce the same proportion of the progeny from that group, rather than (as might otherwise be assumed) a player at higher frequency in a group producing a larger proportion of all progeny produced by the group. This result can be demonstrated by examining the frequency-weighted fitness of any contributor via substituting the expression for the ESS (eqn. 4) into the expression for a player's fitness (eqn. 1), which simplifies to: $p_i\omega_i = (c/b)(1 + b\hat{x}_G)^2$ (Table 2). This expression means that contributors must make unequal contributions to public goods that wipe out the differences in their starting frequencies, which necessarily means that those at higher frequency are investing more (and hence are paying a larger fitness cost) than those at lower frequency. While all contributors end up with equal transmission to the next generation, their transmission is always higher than free-riders (*i.e.* the individual contributors each account for a larger proportion of all progeny produced by members of the group than do each of the free-riders). However, free-riders have higher fitness than contributors because they do not pay for any of the costs of public goods and yet receive the same benefit as contributors.

Kin selection perspective

We can understand the logic of why and how much individuals contribute to public goods by relating the model results to the kin selection perspective. For this we can examine the conditions where a player should contribute toward the production of public goods rather than being a free-rider, which we can express by finding the conditions that satisfy the inequality for when the ESS level of investment is greater than zero:

$$p_i b - c - n(\bar{p} - p_i)b > 0 \quad (5)$$

The first two parts of this expression ($p_i b - c$) are like a simple form of Hamilton's rule (Hamilton, 1964a; Charnov, 1977), whilst the third part $-n(\bar{p} - p_i)b$ captures how a player should adjust their investment strategy in response to the investment by their group-mates. (Formally, Hamilton's rule would describe the quantity of investment at the ESS, which requires examination of its marginal form; see SI Appendix, Section 2.) The expression $p_i b - c$ captures the direct 'profitability' of investment into public goods (in terms of fitness), where each unit of investment results in a reduction in fitness of c and a benefit in terms of $p_i b$. A player's frequency is multiplied by the benefit term because this term scales how much a player's contribution toward collective investment is able to impact that player's own fitness. For example, a player at low frequency in a group is largely incapable of affecting their own fitness through their contribution to public goods, regardless of how much they might invest (because their contribution is diluted through the group in relation to their frequency). This component of equation (5) ($p_i b - c$) directly matches the classic form of Hamilton's rule since the frequency of a player in the group (p_i) also represents its relatedness to the group (which includes itself). The last component of equation (5) ($-n(\bar{p} - p_i)b$) captures the effect of other players' investment on the motivation for the focal player to contribute toward public goods. When a player is at a higher frequency than the average frequency of contributors, this last term will be positive and have no impact on the conditions since the first condition ($p_i b - c >$

0) has to be met for that player to be motivated to invest in the first place. However, if a player is at a frequency that is below the average frequency of contributors, this last term is negative and indicates that such players will have a greater incentive to free-ride on the public goods produced by others (and hence reduce their own contribution). In this way, this term ($-n(\bar{p} - p_i)b$) captures the conflict that arises between players, as they reduce their investment in response to the investment made by other players (*cf.* Queller, 1994; Frank, 1998). Therefore, overall, whether or not a player contributes toward the production of public goods or free-rides depends on a combination of the potential profits of investing from a simple cost-benefit analysis and also the impact of social conflict.

We can also examine the impact of these two factors, the simple cost-benefit analysis that determines the profitability of investing in public goods and the impact of social conflict that can disincentivise contributing, in terms of the quantitative level of investment into public goods that is favoured. The ESS (eqn. 2) can be rearranged into two components that reflect these factors by splitting the collective investment by the group (the coESS) into contributions from the focal player and all other players ($\hat{x}_G = p_i\hat{x}_i + p_{-i}\hat{x}_{-i}$), which are separated into terms by the square brackets:

$$\hat{x}_i = \left[\frac{1}{2}\left(\frac{1}{c} - \frac{1}{p_i b}\right)\right] - \left[\frac{1}{2}\left(\frac{p_{-i}\hat{x}_{-i}}{p_i}\right)\right] \quad (6)$$

The first term in square brackets contains the Hamilton's rule cost-benefit balance that depends solely on a player's frequency or relatedness to the group (p_i). The second term in square brackets contains the effect of other players' investment and, because this bracket is taken away from the first, other player's investment always acts as a 'conflict load' in decreasing the quantity of investment that a player contributes compared to the simple Hamilton's rule. Therefore, although a player's quantitative contributions toward cooperation are not determined by a familiar Hamilton's rule, the quantity of investment is shaped by the

same factors that determine whether or a not a player is a contributor or free-rider.

Group selection perspective

We can also understand the logic of why and how much individuals contribute to public goods by relating the model results to the group selection perspective. For this we can partition a player's fitness into the product of within- (u_i) and between-group (u_G) components ($\omega_i = u_i u_G$). This partitioning is in keeping with group selection models (Haldane, 1932; Wright, 1945; Maynard Smith, 1964; Wilson, 1975a), even though we are not modelling selection on groups per se because we do not consider within- and between-group competition (but rather the ESS that is selected to maximize a player's fitness within any given group context). Nevertheless, we can derive a within-group component of fitness to describe what proportion of the benefits from public goods a player receives compared to their group-mates, whilst the between-group component of fitness describes the overall magnitude of the benefits of public goods to the group. This partitioning yields an expression for within-group fitness:

$$u_i = \frac{p_i(1 - c\hat{x}_i)}{1 - c\hat{x}_G} \tag{7}$$

This indicates that a player has maximal within-group fitness when they do not invest at all (as long as others invest, otherwise all players would have equal fitness). The partitioning also yields an expression for between-group fitness, which is defined as the sum of the fitness of the N players that compose the group ($u_G = \sum_{i=1}^{N} \omega_i$):

$$u_G = (1 - c\hat{x}_G)(1 + b\hat{x}_G) \tag{8}$$

When considering the cost-benefit relationship of public goods for the group as a whole, which is equivalent to considering a group composed of a single player ($N = 1,\ p_i = 1$), the relationship between collective investment and between-group fitness is simply a quadratic function. This function captures the intuitive property of investment into public goods: more investment could potentially yield more benefits but, by sacrificing investment to the production of public goods, players necessarily reduce their ability to benefit from those goods. At the peak of the quadratic, the optimum level of collective investment is:

$$\theta = \frac{b - c}{2bc} \tag{9}$$

Thus, the group has maximal between-group fitness at some intermediate level of investment. In this way, a player's strategy is a compromise that reflects a trade-off between fitness at different levels because within-group fitness is maximised by no investment whilst between-group fitness is maximized by some intermediate level of investment.

The difference between what is good for the player and good for the group can be examined further in a simple analysis of 'alignment' through the effects of parameter variation on the ESS (eqn. 4) and coESS (eqn. 3), where we consider the effect of varying each parameter when all other parameters are kept constant (Table 3). Although some parameters have the same effect on the quantity of investment by a player and quantity of collective investment by the group, such as the benefits (b) and the costs (c) of public goods, others have the opposite effects such as the number (n) and average frequency ($\bar{p}$) of contributors. This captures how players and groups experience a common trade-off at the between-group level over the benefits and costs of public goods, but there is a social dilemma about how players contribute toward public goods arising from the within-group level (because of conflict between players strategies; see eqn. 6). Consequently, increasing a contributor's frequency (p_i) increases their quantity of investment just as increasing the average frequency of contributors ($\bar{p}$) increases their quantity of collective investment, but increasing the average

frequency of contributors ($\bar{p}$) decrease a contributor's quantity of investment. In this way, the conflict between players' strategies can be considered as a conflict between levels (*i.e.* within- and between-group fitness).

Table 3. The effect of increasing each parameter on the level of individual investment ($\hat{x}_i$; see eqn. 4) and collective investment ($\hat{x}_G$; see eqn. 3). The comparison is made keeping all other terms constant.

Parameter	**Individual investment (x_i)**	**Collective investment (x_G)**
Benefit (b)	↑	↑
Cost (c)	↓	↓
Frequency (p_i)	↑	N/A†
Average contributor frequency ($\bar{p}$)	↓	↑
Number of contributors (n)	↑	↓

† This term does not directly appear within the expression for coESS of collective investment (eqn. 3).

Discussion

To understand how individuals should strategically modulate their contributions toward group-beneficial public goods in response to their social context, we analysed how different genetic variants should contribute to public goods in relation to the number and frequencies of other genetic variants in the group. To simplify the discussion of our analyses, we describe the genetic variants to be the 'players' in the game and consider their contributions to public goods as representing an 'investment' they make in terms of the proportion of their potential fitness that they sacrifice to produce public goods (hence their potential fitness gives their total budget they can use to produce public goods). Most intuitively, and confirming the fundamental results of other public goods models (Frank, 1996; Gavrilets, 2015; Johnstone & Rodrigues, 2016; Madgwick *et al.*, 2018), a player should contribute a larger proportion of their fitness budget toward group-beneficial behaviour when they are at a higher frequency in a group. This is because a player at a higher frequency can realise more of the benefits from its own contribution towards public goods, making a larger investment more profitable. For a group composed of multiple players that contribute toward public goods, the level of collective investment made by all players in the group depends on the distribution of frequencies of the players within the group. Somewhat surprisingly, but following other models of facultative cooperation (Boyd & Richerson, 1988; Pepper, 2000; Gardner & Grafen, 2009), we expect a group with more contributing players to contribute less toward group-beneficial behaviour. This outcome is a consequence of each contributor being at a lower frequency in the group, which disincentives each from investing (even in situations where they collectively have the same total frequency as a smaller number of contributors). Thus, within our model of strategic cooperation (*i.e.* not facultative cooperation because we consider quantitative contributions toward public goods), a group of players contribute more toward group-beneficial behaviour when the average frequency of all contributors is larger. Finally, and most surprisingly, we find that all contributors are expected to have equal transmission to the next generation irrespective of differences in their frequencies in the group. When contributors are at different

frequencies in a group, this invariance pattern must reflect compensatory investment, where the contributors at higher frequency sacrifice a larger proportion of their potential fitness to produce public goods than those at lower frequencies, which wipes out any advantage from having a higher frequency. Invariance results have been demonstrated in previous evolutionary models of sex ratio and dispersal (*e.g.* Frank 1987; Rodrigues and Gardner 2016), but our result represents a novel finding among evolutionary models of public goods. Further, whilst analogous results to fitness invariance to frequency have been uncovered in previous economics models of public goods (Warr, 1983; Bergstrom *et al.*, 1986), we demonstrate that invariance can be evolutionarily stable and hence can be biologically relevant in the patterns of strategic cooperation.

Why haven't other evolutionary models of public goods uncovered fitness invariance to frequency? The invariance result is tightly connected to the multiplicative costs and benefits of public goods. Fitness invariance across frequencies requires players to quantitatively adjust their level investment into cooperation to cancel out the quantitative change in their frequencies. Additive costs and benefits lead to facultatively all-or-none cooperation from the threshold within this single fitness effect, as maximal cooperation is favoured when the benefits outweigh the costs (and not otherwise) (Marshall, 2015). Multiplicative costs and benefits lead to quantitively variable level of cooperation from the trade-off between separate fitness effects, so multiplicative fitness effects make fitness invariance possible – and additive models of public goods would not have this possibility. Other similar multiplicative models of public goods (*e.g.* Frank 1995) have failed to derive an informative analytical solution to players' evolutionarily stable strategy which would imply fitness invariance (though it has been speculated; Frank 1996). Our novel method of analysis has enabled fitness invariance to be uncovered, and indeed can be applied to other multiplicative model setups to demonstrate fitness invariance to frequency (see Supplement 3 for an application to Frank 1995).

Fitness invariance to frequency is intriguing because it suggests that players are willing to pay more for public goods up to the level at which they come away with the same

total payoff as all other contributors (in terms of transmission into the next generation), which means that those at lower frequencies sacrifice less of their fitness to produce public goods, while gaining the same benefits. This outcome reflects a property of investing into public goods that is analogous to the snowdrift game (Rapoport & Chammah, 1966; Sugden, 1986). In the classic snowdrift game, a player is favoured to cooperate even when others do not, so an arbitrary asymmetry can cause players to adopt different roles (of contributor and free-rider). In our model of strategic cooperation, the critical asymmetry is not arbitrary, but rather, reflects the fact that players are at different frequencies in a group. Those at a higher frequency have more to gain from each unit they contribute to public goods, and hence have more incentive to contribute. Logically, if players at a higher frequency contribute more than those at a lower frequency, the frequency-weighted fitness (which is their transmission into the next generation) asymmetry between players is diminished, which ultimately leads to an outcome where all contributors make contributions to public goods such that there is no remaining asymmetry. Interestingly, any free-rider always has higher fitness than any contributor since they sacrifice none of their potential fitness to produce public goods, but gain the same benefits as the contributors. Therefore, in groups that contain contributors, players will have negative frequency-dependent fitness, with free-riders having the highest fitness and contributors having fitness in inverse order to their level of investment.

Given that the players are genetic variants that have information about their frequency in the group (such as a greenbeard or kin recognition gene), negative frequency-dependent fitness could lead to the maintenance of polymorphism at the causal locus. As public goods give a fitness advantage to lower frequency players that free-ride on the contributions of others (or otherwise invest less than other contributors), a new mutant that can differentiate itself as a player (presumably by producing a different signal) would potentially have a fitness advantage because of its low frequency. Consequently, even though all genetic variants are expected to play the same strategy (the evolutionarily stable strategy; ESS), negative frequency-dependence arising from the quantitative adjustment of the level of cooperation

within groups may lead to polymorphism by favouring rare genetic variants. In this way, strategic cooperation may explain how variation in social recognition persists, as a potential solution to Crozier's paradox (Crozier, 1986; Rousset & Roze, 2007). Further, although this is not something that we model (as we focus on the patterns of a player's ESS across social contexts), in longer-term evolution, polymorphism may lead to worse outcomes for the production of public goods in groups, which may contain a large number of genetic variants at low frequencies, such that each may lack any incentive to be cooperative.

The ESS pattern of cooperation through public goods is consistent with the expectation from kin selection theory, but our analysis reveals a role for conflict that is not present in the simplest statement of the conditions for cooperation that is captured by Hamilton's rule. Within this analysis, a player's frequency in the group is equated with their relatedness to the group. Although the level at which players contribute to public goods reflects the relatedness-dependent profitability of such contributions (in terms of the cost-benefit analysis expressed in Hamilton's rule), players are expected to reduce their own contribution below the simple cost-benefit analysis when other members of their group are favoured to contribute toward public goods. This outcome captures the role of conflict in cooperation through group-beneficial public goods, where players potentially contribute at a level that is lower than expected under the strict profitability analysis of Hamilton's rule. These same results can be understood from a levels of selection perspective, where contributions to public goods can reduce a player's within-group fitness because they sacrifice a component of their potential fitness by contributing toward public goods, which can leave them with lower relative fitness compared to members of their group who contribute less. As a result, players maximise their within-group fitness by contributing less than others, which can drive down the incentive to contribute. However, contributions to public goods benefit players through their impact on the fitness of the group (since all members of a group gain the same benefits) and therefore the ESS level of investment in public goods will reflect the

disincentivising effect of within-group fitness and the incentivising effect of between-group fitness.

Acknowledgements

This work was supported by funding from the BBSRC to P. G. M. (through the SWBio DTP) and J. B. W. (BB/M01035X/1).

References

Abbot, P., Withgott, J.H. & Moran, N.A. 2001. Genetic conflict and conditional altruism in social aphid colonies. *Proc. Natl. Acad. Sci.* **98**: 12068–12071.

Bergstrom, T., Blume, L. & Varian, H. 1986. On the private provision of public goods. *J. Public Econ.* **29**: 25–49.

Bourke, A.F.G. 2011. *Principles of Social Evolution*. Oxford University Press.

Boyd, R. & Richerson, P.J. 1988. The Evolution of Reciprocity in Sizable Groups. *J. Theor. Biol.* **132**: 337–356.

Charnov, E.L. 1977. An elementary treatment of the genetical theory of kin-selection. *J. Theor. Biol.* **66**: 541–550.

Clutton-Brock, T. 2009. Cooperation between non-kin in animal societies. *Nature* **462**: 51–57.

Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature 373*: 209–216.

Cosmides, L.M. & Tooby, J. 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**: 83–129.

Crozier, R. 1986. Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution (N. Y).* **40**: 1100–1101.

Dionisio, F. & Gordo, I. 2006. The tragedy of the commons, the public good dilemma and the meaning of rivalry and excludability in evolutionary biology. *Evol. Ecol. Res.* **8**: 321–332.

Drescher, K., Nadell, C.D., Stone, H.A., Wingreen, N.S. & Bassler, B.L. 2014. Solutions to the public goods dilemma in bacterial biofilms. *Curr. Biol.* **24**: 50–55.

Ferrari, M., Lindholm, A.K. & König, B. 2016. A reduced propensity to cooperate under enhanced exploitation risk in a social mammal. *Proc. R. Soc. B Biol. Sci.* **283**: 20160068.

Ferrari, M., Lindholm, A.K. & König, B. 2015. The risk of exploitation during communal nursing in house mice, Mus musculus domesticus. *Anim. Behav.* **110**: 133–143.

Foster, K.R. 2004. Diminishing returns in social evolution: The not-so-tragic commons. *J. Evol. Biol.* **17**: 1058–1072.

Fox, J. & Guyer, M. 1978. "Public" Choice and Cooperation in n-Person Prisoner's Dilemma. *J. Conflict Resolut.* **22**: 469–481.

Frank, S.A. 2010. A general model of the public goods dilemma. *J. Evol. Biol.* **23**: 1245–1250.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press.

Frank, S.A. 1987. Variable sex ratio among colonies of ants. *Behav. Ecol. Sociobiol.* **20**: 195–201.

Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520–522.

Frank, S.A. 1996. Policing and group cohesion when resources vary. *Anim. Behav.* **52**: 1163–1169.

Frank, S.A. 2003. Repression of competition and the evolution of cooperation. *Evolution (N. Y).* **57**: 693–705.

Gardner, A. & Grafen, A. 2009. Capturing the superorganism: A formal theory of group adaptation. *J. Evol. Biol.* **22**: 659–671.

Gavrilets, S. 2015. Collective action problem in heterogeneous groups. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**: 20150016.

Gaynor, K.M. & Cords, M. 2012. Antipredator and social monitoring functions of vigilance behaviour in blue monkeys. Anim. Behav. 84: 531–537.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E. & Queller, D.C. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc. Natl. Acad. Sci.* **104**: 8913–7.

Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology* (J. R. Krebs & N. B. Davies, eds), pp. 62–84. Blackwell Science Ltd.

Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024–1027.

Haldane, J.B.S. 1932. *The Causes of Evolution*. Longmans, Green and Co. Limited.

Hamburger, H. 1973. N-person Prisoner's Dilemma. *J. Math. Sociol.* **3**: 27–48.

Hamilton, W.D. 1964. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**: 1–16.

Johnstone, R.A. & Rodrigues, A.M.M. 2016. Cooperation and the common good. *Phil Trans R Soc B* **371**: 20150086.

Kirkpatrick, R. & Gomulkiewicz, M. 1992. Quantitative Genetics and the Evolution of Reaction Norms. *Evolution (N. Y).* **46**: 390–411.

Madgwick, P.G., Belcher, L.J., Stewart, B., Thompson, C.R.L. & Wolf, J.B. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. *Proc. Natl. Acad. Sci.* **115**: E4823–E4832.

Marshall, J.A.R. 2015. *Social Evolution and Inclusive Fitness Theory*. Princeton University Press.

Maynard Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge University Press.

Maynard Smith, J. 1964. Group Selection and Kin Selection. *Nature* **201**: 1145–1147.

Maynard Smith, J. 1974. The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **47**: 209–221.

Parker, G.A. & Maynard Smith, J. 1990. Optimality theory in evolutionary biology. *Nature* **348**: 27-33

Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.

Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.

Rankin, D.J., Bargum, K. & Kokko, H. 2007. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**: 643–651.

Rapoport, A. & Chammah, A.M. 1966. The Game of Chicken. *Am. Behav. Sci.* **10**: 10–28.

Riehl, C. 2013. Evolutionary routes to non-kin cooperative breeding in birds. *Proc. R. Soc. B Biol. Sci.* **280**: 20132245.

Rodrigues, A.M.M. & Gardner, A. 2016. The constant philopater hypothesis: A new life history invariant for dispersal evolution. *J. Evol. Biol.* **29**: 153–166.

Rousset, F. & Roze, D. 2007. Constraints on the Origin and Maintenance of Genetic Kin Recognition. *Evolution (N. Y).* **61**: 2320–2330.

Santema, P. & Clutton-Brock, T. 2013. Meerkat helpers increase sentinel behaviour and bipedal vigilance in the presence of pups. *Anim. Behav.* **85**: 655–661.

Sugden, R. 1986. *The Economics of Rights, Co-operation and Welfare*. Blackwell.

Travisano, M. & Velicer, G.J. 2004. Strategies of microbial cheater control. *Trends Microbiol.* **12**: 72–78.

Trivers, R.L. 1971. The Evolution of Reciprocal Altruism. *Q. Rev. Biol.* **46**: 35–57.

Velicer, G.J., Kroos, L. & Lenski, R.E. 2000. Developmental cheating in the social bacterium Myxococcus xanthus. *Nature* **404**: 598–601.

Warr, P.G. 1983. The private provision of a public good is independent of the distribution of income. *Econ. Lett.* **13**: 207–211.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: R661–R672.

West, S.A., Griffin, A.S., Gardner, A. & Diggle, S.P. 2006. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4**: 597–607.

Wilson, D.S. 1975. A theory of group selection. *Proc. Natl. Acad. Sci.* **72**: 143–146.

Wilson, D.S. 1977. Structured demes and the evolution of group-advantageous traits. *Am. Nat.* **111**: 157–185.

Wilson, E.O. & Hölldobler, B. 1990. *The Ants*. Harvard University Press.

Wright, S. 1945. Tempo and Mode in Evolution : A Critical Review. *Ecology* **26**: 415–419.

Comment 3: Selection for polymorphism

The preceding paper provides a detailed strategic analysis of a generalised form of the Collective Investment game, explaining the factors that shape a player's strategy. Here, with a particular focus on negative frequency-dependent fitness arising from conditional cooperation as a solution to Crozier's paradox, I evaluate how robust these results are likely to be.

A major difference between investment strategy in Paper 3 and natural systems is that players are likely to have informational constraints about their social context (see also Comment 2). As the only variable aspects of a social context that are relevant for a player's strategy are their own frequency and the investment made by others, we would anticipate that these two social variables give two different kinds of informational constraint. Firstly, players may inaccurately assess their own frequency in a group, which should typically reflect error in their assessment of their relatedness to the group (given that the players are genetic variants). Whilst this would undoubtedly change the exact quantitative predictions, strategies are monotonic and so we would still expect that players would invest a more into public goods when they are at higher frequencies. Secondly, players may be unable to assess the composition of a group, which most plausibly means only having information on their own representation in the group instead (*i.e.* their own frequency in or relatedness to the group). Consequently, players may not be able to modulate their investment into public goods to their specific social context. As a result, we expect strategies to reflect a form of bet-hedging that yields the best response to the expected social context. Under these circumstances, the behaviour of players is likely to be maladaptive outside of the social contexts most often experienced in nature. Nonetheless, our model predicts that strategies that have evolved under such bet-hedging would be limited within specific bounds that depend on the expected level of investment being made by others (see Figure 2 in Paper 3). As the partitioning in equation (6) in Paper 4 emphasises, the ESS varies between the high investment scenario where others

are expected to make no investment (which occurs when there are a large number of group-mates which each have too low a frequency to incentivise their own investment) and the low investment scenario where others make the maximal contribution toward public goods (which occurs when there is only one other group-mate).

Generally, the quantitative pattern that players invest proportionally more into public goods when they have a higher frequency is likely to be robust to all constraints (see Paper 2), but the effects of changing the numbers of players in a group and whether there are situations with compensatory investment would depend upon the absence of plausible constraints. Herein lies the value of the properties the game's predictions: given that the quantitative predictions about a player's strategy are likely to be robust even if they have relatively imperfect information about their social context, the other more sensitive predictions about changing the numbers of players in a group and whether there are situations with compensatory investment can be used to interrogate what information players have access to (see Comment 2). As the genetic/phenotypic mechanisms that enable the quantitative adjustment of the level cooperation in different situations remains poorly understood in many of the study systems, these predictions provide an approach to discovering how players gain information about their social context and carry out strategic cooperation.

Because the prediction that players invest proportionally more into public goods when they have a higher frequency is likely to be robust, the finding that conditional cooperation can lead to negative frequency-dependent fitness is also likely to be robust. The generalisation of the Collective Investment game clarifies that the real players of the game would not be strains as we have constructed it in previous presentations of the model for our analysis of *D. discoideum* (in Papers 1 and 2). Instead, the player is a genetic variant that is responsible for deciding whether a biological entity belongs to one player or other, like a greenbeard or kin-recognition allele. A major problem for the theory of genetic recognition mechanisms is the maintenance of the polymorphism that is needed for these systems to work – a problem known as Crozier's paradox. It has long been held that genetic recognition mechanisms controlling

the expression of cooperative behaviour is highly unlikely because there would be a common-type advantage leading one genetic variant to go to fixation and eliminate the polymorphism required for genetic recognition. The common variant gains this advantage because its frequency means that is more likely to interact with other variants that share a recognition signal, and therefore it is simply more likely to engage in cooperation. Models highlighting this problem are based on facultative cooperation, where cooperation switches between non-cooperation and full-cooperation (*e.g.* Crozier, 1986; Rousset & Roze, 2007; Ben-Zion *et al.*, 2018), which is very different from the Collective Investment game where there are continuous levels of cooperation through public goods. At the ESS (*i.e.* when a single frequency-dependent strategy has gone to fixation), conditional cooperators invest more into cooperation when they are at higher frequency, which affords a fitness advantage to lower frequency conditional cooperators. In this way, we would expect the conditional cooperation to lead to diversifying selection which could be a broadly applicable mechanism for the maintenance of polymorphism.

Paper 3 analysed a generalised form of the Collective Investment game and discovered that players (self-signal variants) are expected to have negative frequency-dependent fitness, which Comment 3 has suggested is likely to be broadly applicable whenever there is conditional cooperation. Next, I examine what this might mean for the selection of the mechanism of conditional cooperation across study systems, where there is a significant dearth of evidence – and so I focus on the recent suggestions of conditional cooperation via greenbeard genes. Whilst this may seem like an odd restriction, candidate greenbeard genes are amongst the best studied genetic recognition systems: with a clear candidate gene, a suggested signalling mechanism and a specific social effect. Many of these genes have also been suggested to be kin-recognition genes, and this does not dispute the utility of looking into them as plausible genes that enable conditional cooperation.

Paper 4

This declaration concerns the article entitled:			
Greenbeard genes: theory and reality			
Publication status (tick one) **Draft manuscript** ☐ **Submitted** ☐ **In review** ☐ **Accepted** ☐ **Published** ✓			
Publication details (reference)	**Madgwick, P. G.**, L. J. Belcher, and J. B. Wolf. 2019. Greenbeard genes: theory and reality. Trends in Ecology and Evolution.		
Copyright status (tick the appropriate statement) I hold the copyright for this material ✓ Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Formulation of ideas: 40%: PGM significantly contributed to the concept, literature review and narrative opinion. Design of methodology: N/A Experimental work: N/A Presentation of data in journal format: 50%: PGM predominantly executed the drafting, revising and editing of the manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	PGMadgwick	**Date**	07/11/2019

Data access statement: N/A

Greenbeards genes: theory and reality

Authors: Philip G. Madgwick[1*], Laurence J. Belcher[1], and Jason B. Wolf[1]

Affiliations: [1] Milner Centre for Evolution and Department of Biology & Biochemistry, University of Bath, Bath, BA2 7AY, UK

[*] Correspondence to: pgm29@bath.ac.uk

Abstract

Greenbeard genes were proposed as a cartoonish thought experiment to explain why altruism can be a selfish strategy from the perspective of genes. The likelihood of finding a real greenbeard gene in nature was thought to be remote because they were believed to require a set of improbable properties. Yet, despite this expectation, there is an ongoing explosion in claimed discoveries of greenbeard genes. Bringing together the latest theory and experimental findings, we argue that there is a need to dispose of the cartoon presentation of a greenbeard to refocus their burgeoning empirical study on the more fundamental concept that the thought experiment was designed to illustrate.

Keyword: kin selection, social evolution, altruism, cooperation, gene conflict

Highlights

- The greenbeard concept was conceived to illustrate why genes are always fundamentally ‘selfish’, rather than to provide empirically-testable predictions about the properties of real genes underlying social behaviours.

- Despite their apparent implausibility in nature, in recent years there has been an explosion of claimed examples of greenbeard genes.
- The theory of greenbeard genes has largely recognised the underlying principle that the concept set out to explain, whilst the empirical study of greenbeards has been constrained by inessential features of its cartoonish presentation.
- Kin selection and the greenbeard effect are alternative explanations for the evolution of a social behaviour, which can be experimentally distinguished with appropriate evidence.
- Experimental findings of how greenbeard genes function and evolve both inform and contradict theoretical expectations.

From thought experiment to real gene

To help readers understand why **altruism** (see Glossary) is selfish from the perspective of genes, Dawkins (1976) reformulated a thought experiment originally devised by Hamilton (1964b), where a gene is able to produce a signal (like a 'green beard'), identify that signal in others, and respond by being altruistic towards those individuals. By directing altruism towards individuals that contain copies of itself, these 'greenbeard' genes benefit their own replication, even if paying the cost of such altruism harms the success of other genes in the same genome. The fanciful nature of Dawkins' cartoon scenario caught the imagination of evolutionary biologists, but greenbeard genes were thought to require a series of special properties that render them biologically unrealistic, or at least highly unlikely to be detected (see Table 1) (Dawkins, 1979, 1982; Grafen, 1984; Crozier, 1986; Hamilton, 2001; Grafen, 2006a; West & Gardner, 2013).

Despite appearing fantastical, in recent years there has been an explosion in claimed discoveries of greenbeard genes (Keller & Ross, 1998; Queller *et al.*, 2003; Smukalla *et al.*, 2008; Pathak *et al.*, 2013; Heller *et al.*, 2016; Gruenheit *et al.*, 2017), as well as the

reinterpretation of known genes as greenbeards (Ridley & Grafen, 1981; Haig, 1996, 2013; Gardner & West, 2010; Unterweger & Griffin, 2016; Danka *et al.*, 2017; Gruenheit *et al.*, 2017). This explosion coincides with the application of new methodologies from molecular biology to understand the evolution of **social behaviour**, especially in focusing on cellular interactions and social behaviours in microbes (Haig, 1996; West *et al.*, 2006; Foster, 2009). With such new focus and techniques, recent studies have found evidence that real genes can exhibit the properties that Dawkins identified (*i.e.* signal, receiver, altruism). However, the scope of these empirical advances has often been impeded by researchers losing sight of the fact that these properties were never intended to represent a rigid set of 'necessary and sufficient' criteria for the **greenbeard effect** (Dawkins, 1979, 1982). Rather, they were presented in this abstract way to provide a simple and intuitive illustration of the fundamental phenomenon of genes manipulating a social behaviour to suit their own self-interest (Dawkins, 1976, 1982). This disconnect has led to confusion about how to apply the greenbeard thought experiment to nature (*e.g.* what constitutes evidence), and ultimately how to convert its abstract logic into a useful concept for empirical research.

In contrast to empirical applications, work on the theoretical side has largely recognised the underlying logic of the fundamental principle that Hamilton (1964b) identified and Dawkins (1976) set out to explain (Hamilton, 1964b, 1972, 1975; Ridley & Grafen, 1981; Queller, 2011; Grafen, 1984, 1985; Queller, 1984, 1992; Biernaskie *et al.*, 2011, 2013; Marshall, 2015). Therefore, while empirical work has been constrained by Dawkins' cartoonish setup of the greenbeard thought experiment, theory has continued to expand applications of Hamilton's concept to understand its role in broader evolutionary phenomena. In the process, theoretical advances have shifted expectations about when and where greenbeard genes can arise, how they evolve, and the kinds of situations that can favour their evolution. These advances have even applied the greenbeard concept to shed light on seemingly disparate evolutionary problems: *e.g.* habitat choice (Dawkins, 1979; Pepper &

Smuts, 2002), Müllerian mimicry (Guilford, 1985, 1988), interspecific mutualism (Frank, 1994; Quickfall & Marshall, 2017), sexual selection (Keller, 2002; Faria *et al.*, 2018).

Despite the continued empirical pursuit of greenbeard genes in nature and the concurrent development of related theory, there continues to be a disconnect between the two which hinders progress on both sides. Empirical researchers are still often focused on seeking evidence for the presentation of the greenbeard concept as laid out by Dawkins (1976). Theoretical research often operates at a level of abstraction that provides limited grounding in natural systems. Here, making use of recent advances from empirics and theory, we argue that there is a timely need to completely dispose of the cartoon illustration of a greenbeard presented by Dawkins in order to refocus empirical research on the more fundamental concept originally laid out by Hamilton (1964b), and to guide the further development of greenbeard theory towards models that make predictions that can be tested in empirical research.

Table 1. Arguments for why greenbeards are unlikely to be found in nature.

Argument	For	Against
	Existential arguments: why greenbeards are unlikely to evolve in the first place	
Far-fetched pleiotropy	Greenbeards require three distinct functions (Hamilton, 1964b; Dawkins, 1976, 1982). Such pleiotropy seems highly unlikely because it requires a single locus – whether being a single gene or multiple linked genes – to gain access to the right information and, at the same time, have the ability to modulate a social behaviour (Dawkins, 1976, 1982; Gardner & West, 2010).	Some types of genes can encode proteins with the three functions (*e.g.* genes for cell surface receptors) (Haig, 1996, 2013) or multiple genes can form a greenbeard locus by linkage. But, a locus does not need to directly 'encode' those three functions to be a greenbeard (*e.g.* it could regulate rather than produce the social behaviour) (Hamilton, 1975; Haig, 1996; Pepper & Smuts, 2002).
Functional integration	Greenbeards require a relatively deterministic mapping between the genotype and phenotype (West *et al.*, 2007a; Zhang & Chen, 2016), which is not found in many species, especially not among vertebrates where development introduces significant environmental variation into organisms' phenotypes (which is liable to provide information for kin recognition via the 'armpit effect') (Dawkins, 1982; Hamilton, 2001; West *et al.*, 2007a).	The strongest evidence for genes evolving under the greenbeard effect comes from microbes that have a simple relationship between genotype and phenotype (Gardner & West, 2010). However, we might also expect greenbeards in 'complex' multicellular organisms, governing molecular-level interactions (rather than behaviours that are controlled by the central nervous system), which can make the genotype-phenotype relationship simpler (Haig, 1996, 2013; Springer *et al.*, 2011).
	Detection arguments: why greenbeards should be hard to empirically recognise	
Fixation removes the greenbeard effect	Greenbeards are likely to evolve under strong selection, leading a greenbeard allele to rapidly spread to fixation (Dawkins, 1982; Crozier, 1986). Once at fixation, the allele is no longer detectable as a greenbeard because it would not exhibit a conditional social behaviour (Dawkins, 1982; Queller *et al.*, 2003; Gardner & West, 2010; Biernaskie *et al.*, 2013).	There can be genetic constraints that prevent a single greenbeard allele from reaching fixation, like homozygote lethality (*e.g.* Gp-9) (Hurst & McVean, 1998; Keller & Ross, 1998; Hamilton, 2001) or, more commonly, a rare-type advantage (Grafen, 1990; Jansen & van Baalen, 2006; Biernaskie *et al.*, 2013; Krupp & Taylor, 2015).
Modification, or host-species extinction	When greenbeards are involved in costly social behaviours that reduce individual fitness, a modifier allele at another locus could benefit from silencing the expression of a greenbeard (Alexander & Borgia, 1978; Dawkins, 1982; Rothstein & Barash, 1983). If greenbeard genes were not modified, they might rapidly accumulate and – owing to their deleterious effects on individual fitness – drive their host-species to extinction (Hamilton, 2001).	Although silencing a greenbeard gene can sometimes benefit a modifier, this may not be possible due to pleiotropic constraints (*i.e.* essential functions). Some greenbeard genes can provide a net benefit to the individuals that carry them, potentially making them immune to modification (Ridley & Grafen, 1981). After-all, a greenbeard is selected because its benefits outweigh its costs, so greenbeards are unlikely to cause host-species extinction (Pepper &

		Smuts, 2002; Gardner & West, 2010; Biernaskie *et al.*, 2011).
	Selection arguments: why greenbeards would be not be favoured by selection	
Falsebeard-driven extinction†	The association between the signal-receiver and behaviour functions of the greenbeard is liable to be disrupted by partial modification by genes at other loci, recombination (if the greenbeard is formed by linked genes), or mutation to knock-out the behaviour function, which can produce 'falsebeards' that possess the greenbeard signal but do not engage in a social behaviour. Because the social behaviour is expected to be costly, falsebeards can cheat greenbeards and drive them extinct (Dawkins, 1976, 1982; Ridley & Grafen, 1981; West *et al.*, 2007a; Gardner & West, 2010; Biernaskie *et al.*, 2011, 2013).	In some systems, falsebeards may not be able to evolve due to the signal-receiver directly causing the social behaviour (Haig, 1996). When falsebeards can evolve, greenbeards can persist by rare-type advantage, which leads to signal-receiver polymorphism ('beard colour' variants) (Grafen, 1990; Jansen & van Baalen, 2006; Biernaskie *et al.*, 2013; Krupp & Taylor, 2015). Such signal-receiver polymorphism reduces the advantage of being a falsebeard, which can lead to their extinction (or non-invasion) or a mixed equilibrium with greenbeards and falsebeards (Grafen, 1990; Jansen & van Baalen, 2006).
Stringent take-off conditions	Many types of greenbeards have frequency-dependent invasion conditions (see Box 1). The benefits of being a greenbeard only arise once a greenbeard allele's frequency is above a critical threshold, because the social behaviour is otherwise too costly (Grafen, 2006a; Jansen & van Baalen, 2006; Gardner & West, 2010; Biernaskie *et al.*, 2011, 2013).	This problem affects greenbeards of all types except facultative-helping because all other types of greenbeard express their costly social behaviour when rare, which can also be accentuated by higher signal costliness ('beard cost') (Gardner & West, 2010; Biernaskie *et al.*, 2011, 2013). Population structure can alleviate this condition by giving new mutants locally higher frequency (Gardner & West, 2010; Faria *et al.*, 2018).

† Falsebeard-driven extinction is the primary reason why greenbeards have not been expected to be found in nature since the greenbeard thought experiment was originally conceived.

The fundamental principle in the greenbeard concept

Dawkins' (Dawkins, 1976) presentation of the greenbeard thought experiment was a simple illustration of Hamilton's (Hamilton, 1964b) more general scenario, and so many of the features of cartoonish presentation are inessential. In the general scenario, a greenbeard gene enhances its fitness by modulating the targeted recipients of a social behaviour via an '**assortment factor**' (Pepper & Smuts, 2002), which indicates (or is at least associated with) the presence or absence of the gene within social partners. For example, the greenbeard could create the assortment factor, such as a 'green beard' phenotype, and cause individuals to behave altruistically in response to its presence. But, a greenbeard gene need not create the assortment factor, it only needs to respond toward it (Hamilton, 1975; Dawkins, 1979). For example, the greenbeard could cause individuals to follow a scent to a particular flower species and then act altruistically to those that are on that flower. In this case, the assortment factor would be the act of following that flowers' scent, which increases the likelihood that the recipients of the altruistic behaviour share the greenbeard gene. In this way, Dawkins' (Dawkins, 1976) presentation of the greenbeard thought experiment is but one of the ways which a greenbeard can enhance its own fitness (see Box 1 for further details).

The greenbeard effect is conceptually analogous to other related, but fundamentally different, forms of selection. Hence it is important that we distinguish the properties that make greenbeard genes different from genes shaped by these other forms of selection. Most critically are kin-selected genes, which, like greenbeard genes, also increase their own fitness by modulating the recipients of social behaviour (Hamilton, 1964a; Grafen, 1985; Frank, 1998). However, despite their conceptual similarity, the two types of genes give rise to fundamentally different fitness effects, reflecting the key differences in the process driving their evolution. **Kin selection** relies upon the individuals being affected by the social behaviour sharing the causal gene with the actor due to common ancestry (Dawkins, 1982; Grafen, 1985; Frank, 1998; West & Gardner, 2013). But common ancestry does not ensure that individuals share an allele at a particular locus, it only means that individuals have some

increased probability of sharing alleles at any locus in the genome (which is determined by the degree of **relatedness**). In contrast, greenbeard genes utilise an assortment factor that *specifically* changes the probability of sharing an allele at the greenbeard locus. Hence, we would expect to see elevated relatedness of interactants at the greenbeard locus, while all other loci would show the background level of relatedness (determined by the ancestral relationship of the individuals). Thus, a greenbeard gene can enhance its own fitness in excess of the other genes in the rest of the genome, whilst a kin-selected gene enhances its own fitness alongside the other genes in the rest of the genome. In this way, the greenbeard effect is a form of '**kind selection**' (Queller, 2011), which clarifies that a greenbeard (unlike a kin-selected gene) is a **selfish genetic element**. Like other kind-selected (*e.g.* meiotic drive) genes, a greenbeard can generate genetic conflict with other loci elsewhere in the genome (Biernaskie *et al.*, 2011; Gardner & Úbeda, 2017). However, a greenbeard differs from other selfish genetic elements by enhancing its fitness using an interaction between individuals (*i.e.* social behaviour) rather than within an individual (*e.g.* gamete killing). This difference can have important ramifications for how gene conflict plays out (Ridley & Grafen, 1981; Gardner & West, 2010; Biernaskie *et al.*, 2011), and therefore it is important to differentiate greenbeards from other types of selfish genetic element. Thus, when contrasted with other related phenomena (kin selection and other forms of kind selection), the critical feature of the greenbeard concept is that a gene is selected because it manipulates a social behaviour to enhance its own fitness in excess of other genes in the rest of the genome.

In empirical studies, the distinction between greenbeards and kin-selected genes has historically been ignored because greenbeards were dismissed as biologically unrealistic. Consequently, studies have tended to focus on testing whether or not a social behaviour evolves under **individual selection** or kin selection, rather than finding ways to distinguish between kin and kind selection. Furthermore, because of the challenges inherent in identifying the genes governing social behaviours, empirical research has tended to make assumptions like the 'phenotypic gambit' (Grafen, 1984) that explicitly ignore any genetic conflicts that

could indicate a role for greenbeards. However, recent advances in molecular biology have enabled the first steps towards a greater understanding of the genetics behind social behaviours, permitting greenbeards to become a more easily testable explanation of social behaviour.

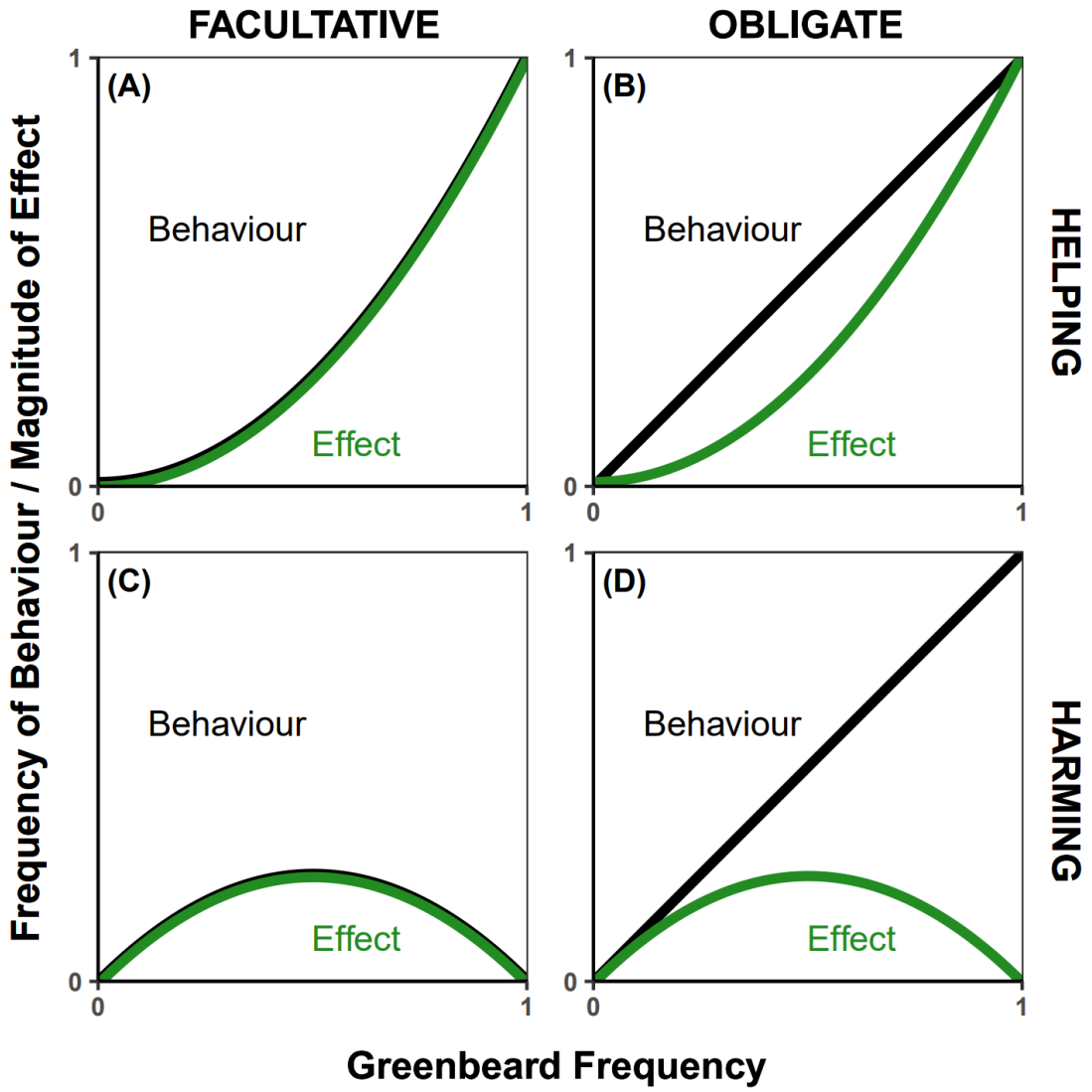
Box 1: Different types of greenbeard genes

Greenbeards manipulate social behaviour to enhance their own fitness in excess of other genes in the rest of the genome using an assortment factor that ensures that they direct the social behaviour towards individuals dependent upon their possession of the greenbeard gene (Hamilton, 1964a; Grafen, 1985; Frank, 1998). The possible mechanisms by which a greenbeard can achieve this outcome are highly variable: using phenotypic or environmental assortment factors, different forms of Hamiltonian social behaviours (altruistic, mutualistic, selfish, and spiteful behaviours; Hamilton, 1964a), targeting individuals with or without the greenbeard gene and so on. Because the possible mechanisms that greenbeards can exploit are so varied, greenbeards are more often grouped into types based on their associated evolutionary dynamics. To this end, two details of their mechanisms are critical, which in combination create four basic types of greenbeard (Gardner & West, 2010) (Box 1 Figure I). This is not to say that outside the abstract neatness of theoretical classifications real genes could not be intermediate types (*e.g.* Unterweger & Griffin, 2016) – as, indeed, this is what we often find (see Table 2).

Greenbeards can be helping or harming, which determines whether their associated social behaviour causes recipients to gain or lose fitness (Gardner & West, 2004b, 2010). A helping greenbeard targets a beneficial effect toward individuals with a copy of itself, and therefore the behaviour increases in frequency as the greenbeard allele increases in frequency (Box 1 Figure IA and IB). In contrast, a harming greenbeard targets a detrimental effect toward individuals without a copy of itself, which occurs at its maximum at intermediate frequency because of the balance between having a large frequency of social partners to both give and receive harm (Figure IC and ID).

Greenbeards can also be facultative or obligate, depending on how actors and recipients engage in the social behaviour (Queller, 1984; Gardner & West, 2010). A facultative greenbeard only pays the cost of performing the social behaviour upon interacting with a social

partner that the social behaviour targets (*i.e.* conditional action), whilst an obligate greenbeard always pays the cost of performing the social behaviour, but the social partner is only affected by it if they are the social behaviour's target (*i.e.* conditional response). Consequently, a facultative greenbeard pays the costs of social behaviour at the rate at which it encounters target social partners (Figure IA and IC), whilst an obligate greenbeard always pays a fixed cost of social behaviour irrespective of whether it finds its target (Figure IB and ID).



Box 1 Figure I. The patterns of frequency-dependence for the four different types of greenbeard (figure redrawn from Gardner & West, 2010). The x-axis is the greenbeard allele's frequency ($0 \leq p \leq 1$) and the y-axis corresponds to either the frequency of the greenbeard-associated social behaviour in the population (black) or the magnitude of the fitness effect for greenbeard carriers (green). For obligate greenbeards (**B**&**D**), the frequency of the behaviour equals the frequency of the greenbeard gene p since the behaviour is always expressed. For facultative greenbeards (**A**&**C**), the frequency of the behaviour is the frequency at which greenbeard carriers interact with either other greenbeard carriers p^2 (facultative-helping) or non-carriers $p(1-p)$ (facultative-harming). Facultative greenbeards (**A**&**C**) receive the fitness effect of the behaviour at the same frequency as they exhibit the behaviour, since the behaviour is only exhibited in the presence of the target recipients (either carriers or non-

carriers). Obligate greenbeards (**B**&**D**) receive the fitness effect at the frequency that carriers interact with recipients, which equals p^2 for helping greenbeards and $p(1-p)$ for harming greenbeards.

Evidence that a gene is a greenbeard

Why do we want to describe real genes as greenbeards? The greenbeard thought experiment, was never intended as an empirically-useful concept, but describing a gene as a greenbeard can be a useful working hypothesis for understanding its evolution (*e.g.* generating revealing predictions). A greenbeard is a selfish genetic element, enhancing its own fitness in excess of other genes in the rest of the genome, and so a greenbeard hypothesis suggests 'who benefits' from the gene's function, which could explain or predict unusual properties. To demonstrate that a gene is a greenbeard, there would need to be conclusive evidence that the gene has evolved (or is evolving) because of its greenbeard effect. It would obviously be necessary to show that there are allele-specific outcomes for social interactions, but exactly what represents 'conclusive evidence' is an open question that we believe is best settled within the constraints of an empirical system (see also Outstanding Questions).

We believe that all empirical systems currently fall short of 'conclusive evidence', but we propose that there is a pivotal piece of (often absent) evidence required to constructively hypothesise that a gene is a greenbeard. There must be evidence that the assortment factor is – at least, in part – independent of common ancestry (see Box 2). We consider kin selection to be the 'null hypothesis' for why a gene would modulate social behaviour, and so kind selection has the onus of proof. To this end, patterns associated with a candidate gene must be examined within natural settings, because the assortment factor may have a correlation with the genome-wide probability of sharing an allele in nature that is absent in artificial laboratory settings. For example, a gene for helping your neighbours could evolve under kin or kind selection depending on the cause of population structure. For kin selection, limited dispersal could mean that neighbours all share a common ancestor and thereby have the same probability of sharing any gene in the genome (Hamilton, 1964a). For kind selection, if an allele causes individuals to congregate in the same habitat, then neighbours would only have an elevated probability of sharing the greenbeard allele (Dawkins, 1979) and some background level of relatedness at the rest of the genome (which will depend on various

factors, like recombination rates and population viscosity). In a laboratory setting, the population structure may not mirror natural settings and consequently may incorrectly suggest that one or other driver is at work.

Preliminary findings about real genes

Numerous greenbeard genes have been identified across a broad range of biological systems and modes of action: *e.g.* bacteriocins (Gardner & West, 2004b, 2010; Biernaskie *et al.*, 2013), cell-binding proteins (Haig, 1996), contact-dependent inhibition factors (Unterweger & Griffin, 2016; Danka *et al.*, 2017), quorum-sensing pherotypes (Pollak *et al.*, 2016; Ben-Zion *et al.*, 2018), imprinted RNAs (Haig, 2013). However, empirical studies of greenbeards are still often preliminary, with few directly demonstrating that the gene has evolved by kind rather than kin selection in natural settings (Table 2). Nevertheless, these empirical studies can hint at the features of real greenbeards that are underappreciated or contrary to the expectations of current theory (see also Box 2).

To understand the nature of empirical evidence supporting the hypothesis that a gene is a greenbeard, we can first consider a cautionary example where new data has drastically altered the understanding of a proposed greenbeard gene. A queen-killing phenotype was observed in the fire ant *Solenopsis invicta*, which was explained by allelic variation at the *gp-9* locus that encodes an odour-binding protein (Keller & Ross, 1998; Trible & Ross, 2015). The greenbeard was suggested to be facultative-harming (Hurst & McVean, 1998), but subsequently the *gp-9* locus has been located on a social chromosome (*Sb*), which is a large linkage group (or supergene) containing over 600 genes that act as a single greenbeard locus (Wang *et al.*, 2013; Pracana *et al.*, 2017). Such linkage at a greenbeard locus doesn't appear unusual; similar social chromosomes have also been discovered in other ant species (Huang & Wang, 2014; Purcell *et al.*, 2014) and many other candidate greenbeard systems involve large linkage groups (Linksvayer *et al.*, 2013; Thompson & Jiggins, 2014). But it does caution

the ascription of properties to candidate greenbeards as, for example, *Sb* appears responsible for numerous social traits other than queen killing, including polygyny (*i.e.* forming nests with multiple queens). As such, the initial characterisation of the greenbeard was incomplete because it did not recognise *Sb*'s role in both helping (polygyny) and harming (queen-killing) social behaviours. Consequently, out of caution, we restrict our discussion to systems where the putative greenbeard genes that have received the most attention, whilst accepting that all the findings are potentially subject to changes in evidence.

The ever-increasing array of examples of greenbeard genes suggests that theory has underappreciated the relationship between allorecognition (*i.e.* detecting self from nonself) and the greenbeard effect. Although vertebrate immune genes have been discussed as greenbeards (Haig, 1997), it is likely that these genes are predominantly selected for asocial reasons (*e.g.* killing parasites) that are under individual selection. However, allorecognition also applies to organisms that interact with conspecifics in ways that allows allorecognition systems to also serve a role in the genetic recognition of potential social partners, such as in marine invertebrates and bacterial colonies (Grafen, 1990; Crampton & Hurst, 1994). Examples of candidate greenbeards broadly fall into the categories of genes governing aggregation for a particular cooperative endeavour and fusion for longer-term cooperation (Table 2). Fusion often involves rejecting or killing individuals that do not possess the right recognition signal, whether these be migrants or mutants. Therefore, in both cases, greenbeards have a functional role in 'privatising' a group's resources for the exclusive use of those that share the same greenbeard allele. Consequently, greenbeards explain why detecting self from nonself is useful – because avoiding interactions with social partners that are unlikely to cooperate (or, worse, cause harm) (Buss, 1982) enables the causal genes to help others that share the gene and harm those that do not.

As examples of greenbeard genes continue to accumulate, they suggest that greenbeard genes often possess two apparently puzzling properties. Firstly, examples of real greenbeard genes show that they often possess multiple 'colour' variants, which is only

theoretically anticipated when each colour variant has its own '**falsebeard**' cheater (Grafen, 1990; Biernaskie *et al.*, 2013). In general, greenbeard genes are expected to exhibit a common-type advantage against one another leading to monomorphism (Crozier, 1986; Grafen, 1990). However, there is little evidence of any falsebeard alleles in any of the current examples. This empirical observation raises the theoretical challenge of explaining why greenbeard colour polymorphism is maintained. Secondly, many of the examples of greenbeard genes involve multiple (most often a pair of) tightly-linked genes and, in some cases, there are multiple unlinked greenbeard genes involved in the conditional expression of the social behaviour. For example, successful outer-membrane fusion in the social bacteria *Myxococcus xanthus* is reliant upon cells having matching *traA/B* alleles for incipient cell fusion, but fusion only persists if cells have matching alleles at the unlinked *sitAI* locus (Vassallo *et al.*, 2017). To the best of our knowledge, no theoretical analysis has ever examined the co-evolution of two or more greenbeard genes at different loci, presumably because one greenbeard gene was thought implausible enough. Overall, preliminary empirical findings suggest that theoretical work has yet to fully characterise the properties of greenbeard genes, and emphasises an increasing need for theoretical and empirical work to study real greenbeard genes in unison.

Table 2. Greenbeard loci that have received the most attention in the literature, grouped by their mode of action in governing aggregation or fusion.

Gene and Species	Signal-Receiver †	Social Behaviour ‡
	Aggregation: genes governing which individuals come together	
csA Social Amoeba (*Dictyostelium discoideum*) (Queller *et al.*, 2003)	**Causal monomorphic.** The *csA* locus is responsible for the production of the gp80 protein that is necessary for cells to bind one another during streaming, which can be shown in the lab using knock-outs (Queller *et al.*, 2003). In nature, strains do not show allelic variation at the *csA* locus, and so it is not clear whether or not the origin and maintenance of this gene is because of a greenbeard effect.	**Facultative helping.** As adhesion itself is the social behaviour, there is no clear distinction between a 'signal-receiver' and the 'social behaviour', as these two aspects are pleiotropic functions of the same physical act of cell-binding (Haig, 1996; Queller *et al.*, 2003). The *csA* gene does not control any downstream social behaviours like fruiting-body formation, though it could clearly influence the identity of social partners if it were polymorphic.
flo1 Budding Yeast (*Saccharomyces cerevisiae*) (Smukalla *et al.*, 2008)	**Causal monomorphic.** FLO1 is a cell membrane protein that binds cell wall carbohydrates (Smukalla *et al.*, 2008). Variation in the number of 100bp repeats determines binding affinity, where genes with fewer repeats produce FLO1 proteins with stronger binding affinities (Verstrepen *et al.*, 2005). However, this variation does not produce different beard 'colours'. Rather, there is one greenbeard colour (dependent on whether or not cells have a function copy, *flo1+*), and various different 'shades of green' that determine how strongly cells bind one another (Gruenheit *et al.*, 2017).	**Facultative helping.** Cells that bind together are protected from external stresses by reducing contact with the environment through aggregation (Smukalla *et al.*, 2008). There is no clear distinction between a 'signal-receiver' and the 'social behaviour' itself, as these two functions are pleiotropic functions of the same physical act of aggregation.
	Fusion: genes governing which individuals successfully merge together	
gp-9 (or *Sb*) Fire Ant (*Solenopsis invicta*) (Keller & Ross, 1998)	**Informative polymorphic.** The *gp-9* locus produces an odour-binding protein that, along with 9 of 24 other odour-binding genes, are located on the *Sb* social chromosome. *Sb* contains over 600 genes, which have become tightly-linked into a supergene via chromosomal inversion (Wang *et al.*, 2013; Pracana *et al.*, 2017). Unsaturated cuticular hydrocarbons are likely to provide the signal (Trible & Ross, 2015). Genetic variants of the *b* allele (*b′* alleles) have never been shown to naturally or	**Facultative helping/harming**. *BB* homozygote queens are killed by *b*-carrying workers, but *bb* queens die prematurely for developmental reasons, leading to stable behavioural variation (Keller & Ross, 1998; Trible & Ross, 2015). The *gp-9* locus has previously been described as a facultative-harming greenbeard (Hurst & McVean, 1998; Gardner & West, 2010), but *Sb* also controls polygynous nest formation by mated queens returning to their mother's nest (Pracana *et al.*, 2017). Thus, the allele must also be facultative-helping

	experimentally co-occur in the same colony.	because of its pleiotropic effect on social behaviours, which provide the social conditions for queen-killing to become advantageous.
traA/traB Social Bacteria (*Myxococcus xanthus*) (Pathak *et al.*, 2013)	**Informative polymorphic.** The *tra* genes are tightly linked and highly polymorphic, encoding cell-surface receptors that bind to each other (Pathak *et al.*, 2013). There are >60 major recognition groups based on *traA/B* similarity (Cao *et al.*, 2019). Switching a single residue in *traA* is sufficient to change the recognition group of a strain (Cao & Wall, 2017).	**Facultative helping.** Matching at *traA/B* leads to fusion and outer membrane exchange. The exact reasons for this are unknown, but the transfer of lipids and proteins has several potential benefits, including coordinating social interactions and cell repair (Pathak *et al.*, 2013).
sitAI1/2/3 Social Bacteria (*Myxococcus xanthus*) (Vassallo *et al.*, 2017)	**Causal polymorphic.** The *sitAI* genes are polymorphic toxin-antitoxin pairs that are transferred between cells matching at the *tra* locus (Vassallo *et al.*, 2017). In natural conditions, incompatibility between strains often occurs independently of *tra* similarity, and cannot be fully explained by *sitAI* (Wielgoss *et al.*, 2018).	**Obligate harming.** Matching at *traA/B* leads to fusion and outer membrane exchange. After initial fusion, sitA toxins can be transferred to social partners that are killed if they lack the matching sitI immunity protein (Vassallo *et al.*, 2017).
tgrB1/tgrC1 Social Amoeba (*Dictyostelium discoideum*) (Gruenheit *et al.*, 2017)	**Informative polymorphic.** The *tgr* genes are tightly-linked (Benabentos *et al.*, 2009) and are amongst the most diverse genes in the genome (Gruenheit *et al.*, 2017). Tgr proteins are cell surface proteins, where TgrB1 acts as a receptor to bind the TgrC1 ligand (Benabentos *et al.*, 2009; Hirose *et al.*, 2011). Reciprocal transplantation of variant sequences of the *tgr* genes demonstrates that matching *tgr* alleles causes a change in social behaviour due to the act of successful binding, which has downstream consequences for multicellular fruiting-body development.	**Facultative helping.** Because chimeric aggregations can be costly, strains may segregate out and develop a separate fruiting body (Benabentos *et al.*, 2009; Hirose *et al.*, 2011; Gruenheit *et al.*, 2017). Thus, successful aggregation is reliant upon Tgr-binding, as segregation is correlated to relatedness at *tgr* rather than genome-wide relatedness (Gruenheit *et al.*, 2017). Strains are known to exhibit partner-specific social behaviour after initial aggregation in the process of fruiting-body development (Hirose *et al.*, 2011), which is not fully explained by Tgr-binding.
doc1/doc2/doc3 Ascomycete Fungi (*Neurospora crassa*) (Heller *et al.*, 2016)	**Informative polymorphic.** The polymorphic *doc1/2/3* are linked genes that produce cell-surface receptors that predict whether or not strains will successfully fuse together (Heller *et al.*, 2016). Reciprocal allele transfers demonstrate that *doc1* and *doc2* are necessary and sufficient for defining which individuals fuse. There is, however, no evidence that *doc1*/2/3 has any influence on the downstream social benefits of somatic fusion.	**Facultative helping.** Somatic fusion of populations mediated by *doc1/2/3* allows many possible benefits including a reproductive division of labour and benefits from sharing organelles and nutritional resources (Richard *et al.*, 2012). Further, as fusion only occurs with matching *doc* alleles, *doc* is likely to regulate fusion to share benefits (Heller *et al.*, 2016).

fuhc-sec/tm Golden Star Tunicate (*Botryllus schlosseri*) (Gruenheit *et al.*, 2017)	**Informative polymorphic.** The *fuhc* locus contains two linked genes that produce a secreted and a transmembrane protein (Nydam *et al.*, 2013). A match at one or both alleles is required for fusion to occur, otherwise a rejection reaction will occur (Scofield *et al.*, 1982). Although the genes are highly polymorphic (De Tomaso, 2018), lower polymorphism correlates with the probability of successful fusion (Nydam *et al.*, 2013). The exact function of two genes is, however, unclear, as *fuhc* matching has not been shown to be sufficient for successful fusion by allele replacement experiments.	**Facultative helping/harming.** The joining of tissues during fusion has several potential benefits, many of which likely relate to the survival and fecundity benefits of larger colony size (De Tomaso, 2018) or the sharing of public goods (Scofield *et al.*, 1982). If *fuhc* alleles don't match then toxic rejection occurs, which can involve damage and death to cells and tissues (De Tomaso, 2018). The degree of rejection correlates with genetic dissimilarity at the *fuhc* locus (Scofield *et al.*, 1982), suggesting that fusion has an obligate-harming side when a match doesn't occur.
alr1/alr2 Saltwater Hydra (*Hydractinia symbiolongicarpus*) (-)	**Informative polymorphic.** The polymorphic *alr1/2* locus encodes transmembrane proteins that bind to each other. If colonies share both alleles, they will fuse permanently (Cadavid *et al.*, 2004). *alr1/2* variation predicts successful fusion in wild and lab strains (Nicotra *et al.*, 2009) and *in vitro* experiments demonstrate highly specific Alr binding (Karadge *et al.*, 2015).	**Facultative helping.** Fusion of colonies mediated by *alr1/2* has several potential benefits, such as increased colony size aiding strong spatial competition (Karadge *et al.*, 2015). Successful fusion may also help to avoid germline parasitism, given that some cells can differentiate into germ cells throughout the lifecycle.

† *Causal vs informative: a causal gene encodes both the signal-receiver component and the social behaviour itself, because the two are one and the same. For example, the* flo1 *locus houses the gene for a cell surface receptor that binds carbohydrates on the surface of other cells, thereby bringing together cells with* flo1 *alleles (where aggregation itself is the social behaviour because it protects cells from environment stressors). An informative gene only encodes the signal-receiver component, which acts to regulate the downstream social behaviour that is itself encoded by other genes. For example, the* Sb *genes do not directly encode all the proteins involved within the process of polygynous nest formation, but they do encode the signal-receiver proteins that set a whole chain of interactions in motion between a great many proteins leading to these outcomes.*

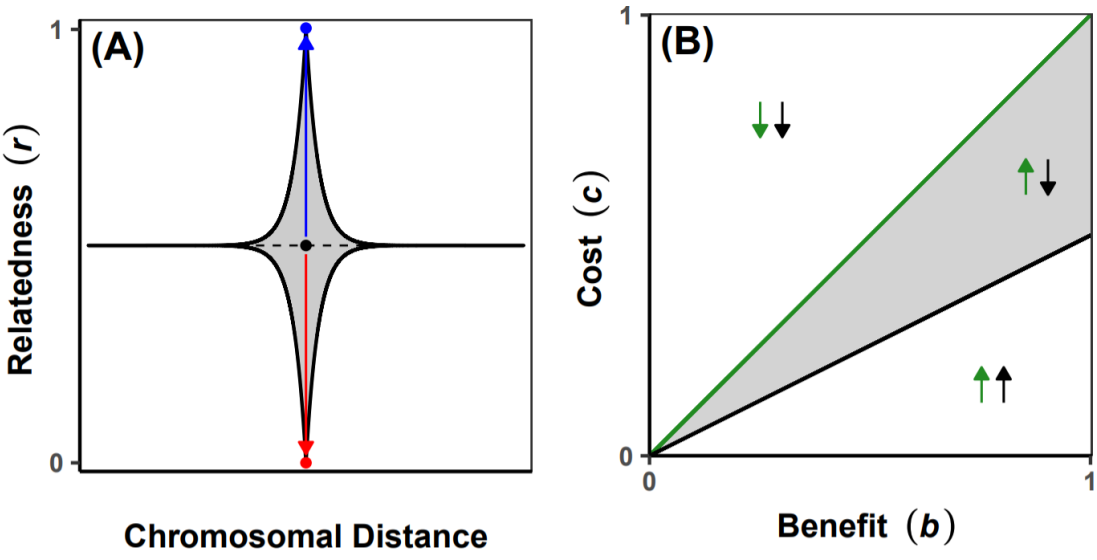
‡ *Monomorphic vs polymorphic: there is only a single greenbeard 'colour' variant, vs multiple colour variants. Colour variation reflects different signal-receiver forms only, rather than different social behaviours in response to signalling.*

Box 2: Are greenbeard genes involved with simple or sophisticated social traits?

Theory has suggested that greenbeards should be associated with simple traits, whilst kin-selected genes would be associated with more complex traits (Grafen, 1985, 2006b; Frank, 1998; West & Gardner, 2013). There are indeed many examples of greenbeards that do little more than produce a protein that directly causes individuals with that protein to aggregate together (see Table 2; *e.g. flo1*, Smukalla *et al.*, 2008). However, suggesting that greenbeards are *only* associated with simplistic traits likely reflects the original scenario envisioned by Dawkins, where the greenbeard gene directly produces a behaviour, which is not a necessary feature of real greenbeard genes. To be favoured by selection, a real greenbeard gene simply needs to be able to modulate social traits in a way that gives them an advantage by ensuring that the social behaviour's recipients share (or do not share) the greenbeard. As such, greenbeards can be 'informative' master regulators – directing a social behaviour toward some recipients rather than others, instead of producing the social behaviour itself (Hamilton, 1964b). In this role, greenbeard genes can be involved with social traits that are just as 'sophisticated' as those associated with kin-selected genes, which is supported by empirical findings because many examples of greenbeards involve fusion (see Table 2; *e.g. doc*, Heller *et al.*, 2016) – a complex social behaviour involving the concerted action of many genes.

Greenbeard genes and kin-selected genes are both favoured because they preferentially affect the fitness of other carriers of the gene, and hence they may often work together in producing the same social trait. However, because they are expected to show a different relationship to social partners, they can also conflict with each other about the social trait's optimum (Biernaskie *et al.*, 2011; Gardner & Úbeda, 2017). This difference arises because the greenbeard's relatedness to social partners depends on its assortment factor, whilst all other (non-greenbeard) genes' relatedness depends on the ancestral relationship between social partners. Consequently, greenbeard genes have greater certainty of sharing (or not) a particular allele with social partners, so they can benefit from more marginal gains than can kin-selected genes (Box 2 Figure I). Although this creates the potential for gene conflict

(Gardner & Úbeda, 2017), a kin-selected gene may be unable to prevent a greenbeard's 'corruption' because of constraints (*e.g.* 'causal' greenbeards; see Table 1) and so the potential conflict may not precipitate into actual conflict (*sensu* Ratnieks *et al.*, 2006). Thus, there is no reason why greenbeards cannot be involved with social traits that are equally as sophisticated as (or indeed may involve other) kin-selected genes.



Box 2 Figure I. The potential for conflict between a greenbeard and kin-selected gene (which acts as a fully-suppressing modifier) that govern the expression of altruism across social trait values. Potential conflict is assessed using a simple Hamilton's rule $rb - c > 0$, where b is a measure of benefits to recipients, c the costs to the actor r gives the relatedness between interacting individuals. The greenbeard gene leads to assortment such that actors interact with recipients that share the greenbeard ($r = 1$), whilst the kin-selected gene leads actors to interact with recipients that are siblings ($r = ½$). For genes along a chromosome (inspired by Grafen, 1985); **A**), the greenbeard (and any linked gene) has greater certainty as to whether two individuals share the greenbeard ($r = 1$; blue) or not ($r = 0$; red) compared to other loci in the genome ($r = ½$; black). Across benefit and cost parameters (inspired by Gardner & Úbeda, 2017); **B**), the greenbeard gene is favoured when $b - c > 0$ (below the green line; indicated by green arrows) whilst the kin-selected gene is favoured when $½b - c > 0$ (below the black line; indicated by black arrows). Under conditions where a greenbeard is favoured whilst the kin-selected gene is not (grey-shading; as shown by green and black arrows pointing opposite directions), there is potential for conflict as a kin-selected gene would be favoured to fully-supress the greenbeard gene.

Concluding remarks and future perspectives

The greenbeard concept was intended as a thought experiment, and it is immensely surprising that experimental research suggests that these genes may actually exist. However, we believe that the future direction of these important empirical studies into candidate greenbeard genes should not get bogged down in trying to interpret real genes through the lens of the thought experiment laid out by Dawkins (1976). Instead, experimental work should focus on the more fundamental underlying concept from Hamilton (1964b), which Dawkins' cartoon was designed to illustrate. With the benefit of hindsight, Dawkins' suggestion of a cartoon scenario of a greenbeard gene is in fact amongst the most unlikely forms of greenbeard, in supposing a pleiotropic gene governing multiple aspects of cognitive behaviour in humans. Although greenbeards are possible in humans, they are more likely to mediate social behaviours via interactions played out at the molecular level rather than the cognitive processes of the central nervous system (see Table 1). Furthermore, the last major review of greenbeard genes (Gardner & West, 2010) highlighted how Dawkins formulation of the greenbeard thought experiment was but one of four possible types of greenbeard gene (whilst also dealing with arising misconceptions). We have emphasised an even broader range of types of greenbeard (informative vs causal), alongside a persistent disconnect between theoretical work that recognised the underlying logic of the fundamental principle originally identified by Hamilton (1964b) and experimental research that has been overly-wedded to features of Dawkins' (1976) presentation of that fundamental principle. Therefore, we propose that there is a need to dispose of the cartoon illustration of a greenbeard presented by Dawkins which can help to align theoretical and experimental research to study greenbeards in unison.

Although there is only preliminary evidence that any gene is a greenbeard, Dawkins cartoon can begin to be replaced with the diversity of genes that capture Hamilton's more fundamental principle – grounding the greenbeard concept within genuine expectations for empirical examples. We are only just beginning to understand the possible functions of greenbeard genes, with preliminary evidence suggesting a common role within allorecognition

in establishing a group of cooperative social partners. For the most part, greenbeards do not appear to be involved with the production of the social behaviour, but instead they act as a master regulator in governing which individuals the social behaviour targets. Within this capacity, greenbeards have a number of genetic properties, several of which are hard to explain from current theory, such as polymorphism of colour variants (rather than different falsebeards).

The study of greenbeards is only in its infancy, but it clear that empirical work has established the plausibility of finding greenbeard genes in nature, and now we need to investigate their importance (see Outstanding Questions). This involves empirically testing kin and kind selection as alternative hypotheses for the evolution of social behaviour, which necessarily involves a greater understanding of the genetics behind social behaviours in order to investigate any candidate greenbeard gene's role. We tentatively suggest that greenbeard genes may be far more common than previously anticipated, contrary to arguments against their biological relevance (Table 1), because species across the tree of life encounter the problem of allorecognition which (from empirical research) appears to be a setting where greenbeards evolve. However, establishing how and why this scenario favours greenbeards remains an open question for theoretical research.

Outstanding questions

At present, there are numerous suggestions for genes that may be greenbeards (or indeed may be kin-selected) and many genes that are currently thought to be kin-selected may turn out to be greenbeards. For empirical research, there is a great need to amass further evidence for both new and existing candidate greenbeards, especially in marrying evidence of plausibility that a gene operates like a greenbeard with evidence that rules out other explanations (especially kin selection). There are some groups of organisms (*e.g.* plants) and types of behaviour (*e.g.* mutualism) where greenbeard genes have yet to be proposed, and it would be interesting to

uncover whether this represents a lack of data or a genuine absence. When approaching any study system, we think there are three critical points to establish if the greenbeard effect is to be constructively suggested to have any role:

1. For a conditionally expressed social behaviour, what exactly is the assortment factor? Are individuals responding to the presence or absence of that factor?
2. Given the recognition of an assortment factor, what exactly is the social behaviour? Have we correctly identified all downstream effects and consequences of assortment?
3. Does the assortment factor correlate with common ancestry in natural settings? Why doesn't kin selection explain the observed patterns of the social behaviour?

For theoretical research, there is a need to seek greater empirical grounding. Within the 'Preliminary findings about real greenbeard genes' section, we identified some key puzzles:

4. How is greenbeard colour polymorphism maintained? And why don't falsebeard cheaters invade or be maintained?
5. How do greenbeards at different loci coevolve? How does relatedness between individuals impact genetic conflict arising from greenbeard genes?
6. What factors drive the linkage of greenbeards from different loci? And when do we expect the evolution of greenbeard supergenes?

Glossary

Social behaviour: any interaction between individuals of the same species. Although the term is sometimes used more broadly to include interspecific interactions, we use the traditional and more restrictive definition that requires interactants to be conspecifics, which critically means that that social partners may be genetically related.

Altruism: social behaviour that is costly to the actor whilst benefitting a recipient.

Individual selection: the form of selection that favours genes because they enhance the fitness of the individual that carries them.

Kin selection: the form of selection that favours a gene with a given social behaviour because of its effect on others that share the gene based on the genealogical relationship between social partners.

Relatedness: the probability that two individuals share the same alleles over and above the random expectation.

Kind selection: the form of selection that favours a selfish genetic element because of a conditional fitness effect, depending on the presence or absence of that specific gene among interactants.

Selfish genetic element: a gene produces a function that enhances its own fitness in excess of the other genes in the rest of the genome.

Greenbeard effect: a form of kind selection, whereby a 'greenbeard gene' enhances its own fitness in excess of the other genes in the rest of the genome by modulating a social behaviour that affects the fitness of others that share the greenbeard gene because of an assortment factor.

Assortment factor: a distinguishing feature that correlates with the presence or absence of a gene within social partners, thereby assorting individuals into those that share the gene and those that do not.

Falsebeard: an allele competing at the same locus as a greenbeard allele that appears to have the greenbeard phenotype but does not perform the associated social behaviour (as a cheater).

Acknowledgements

We would like to thank David Haig and Chris Thompson for helpful discussions, as well as Andy Gardner and three anonymous referees for constructive comments. We would also like to thank our funders: SWBio DTP for PGM, NERC GW4+ for LJB. and BBSRC for JBW (BB/M01035X/1). This work was partially completed as part of a fellowship from the Wisshenshaftkolleg zu Berlin for JBW (and visitation from PGM).

References

Alexander, R. & Borgia, G. 1978. Group Selection, Altruism, and the Levels of Organization of Life. *Annu. Rev. Ecol. Syst.* **9**: 449–474.

Ben-Zion, I., Pollak, S. & Eldar, A. 2018. Clonality and non-linearity drive facultative-cooperation allele diversity. *ISME J.* **13**: 824–835.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E.A., et al. 2009. Polymorphic members of the lag gene family mediate kin discrimination in Dictyostelium. *Curr. Biol.* **19**: 567–572.

Biernaskie, J.M., Gardner, A. & West, S.A. 2013. Multicoloured greenbeards, bacteriocin diversity and the rock-paper-scissors game. *J. Evol. Biol.* **26**: 2081–2094.

Biernaskie, J.M., West, S.A. & Gardner, A. 2011. Are Greenbeards Intragenomic Outlaws? *Evolution (N. Y).* **65**: 2729–2742.

Buss, L.W. 1982. Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc. Natl. Acad. Sci.* **79**: 5337–5341.

Cadavid, L.F., Powell, A.E., Nicotra, M.L., Moreno, M. & Buss, L.W. 2004. An invertebrate histocompatibility complex. *Genetics* **167**: 357–365.

Cao, P. & Wall, D. 2017. Self-identity reprogrammed by a single residue switch in a cell surface receptor of a social bacterium. *Proc. Natl. Acad. Sci.* **114**: 3732–3737.

Cao, P., Wei, X., Awal, R.P., Müller, R. & Wall, D. 2019. A Highly Polymorphic Receptor Governs Many Distinct Self-Recognition Types within the Myxococcales Order. *MBio* **10**: 1–15.

Crampton, W.G. & Hurst, L.D. 1994. True kin recognition, in the form of somatic incompatibility, has multiple independent origins. *Anim. Behav.* **47**: 230–234.

Crozier, R. 1986. Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution (N. Y).* **40**: 1100–1101.

Danka, E.S., Garcia, E.C. & Cotter, P.A. 2017. Are CDI Systems Multicolored, Facultative, Helping Greenbeards? *Trends Microbiol.* **25**: 391–401.

Dawkins, R. 1982. *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford University Press.

Dawkins, R. 1976. *The Selfish Gene*. Oxford University Press.

Dawkins, R. 1979. Twelve misunderstandings of kin selection. *Z. Tierpsychol.* **51**: 184–200.

De Tomaso, A.W. 2018. Allorecognition and Stem Cell Parasitism: A Tale of Competition, Selfish Genes and Greenbeards in a Basal Chordate. In: *Origin and Evolution of Biodiversity* (P. Pontarotti, ed), pp. 131–142. Springer International Publishing.

Faria, G.S., Varela, S.A.M. & Gardner, A. 2018. The relation between R . A . Fisher's sexy-son hypothesis and W . D . Hamilton's greenbeard. *Evol. Lett.* **2**: 190–200.

Foster, K.R. 2009. A defense of sociobiology. *Cold Spring Harb. Symp. Quant. Biol.* **74**: 403–18.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press.

Frank, S.A. 1994. Genetics of mutualism: the evolution of altruism between species. *J. Theor. Biol.* **170**: 393–400.

Gardner, A. & Úbeda, F. 2017. The meaning of intragenomic conflict. *Nat. Ecol. Evol.* **1**: 1807–1815.

Gardner, A. & West, S.A. 2010. Greenbeards. *Evolution (N. Y).* **64**: 25–38.

Gardner, A. & West, S.A. 2004. Spite and the scale of competition. *J. Evol. Biol.* **17**: 1195–1203.

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surv. Evol. Biol.* **2**: 28–89.

Grafen, A. 1990. Do animals really recognize kin? *Anim. Behav.* **39**: 42–54.

Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology* (J. R. Krebs & N. B. Davies, eds), pp. 62–84. Blackwell Science Ltd.

Grafen, A. 2006a. Optimization of inclusive fitness. *J. Theor. Biol.* **238**: 541–563.

Grafen, A. 2006b. Various remarks on Lehmann and Keller's article. *J. Evol. Biol.* **19**: 1397–1399.

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J., Wolf, J. & Thomposon, C. 2017. A polychromatic "greenbeard" locus determines patterns of cooperation in a social amoeba. *Nat. Commun.* **14171**: 1–9.

Guilford, T. 1985. Is kin selection involved in the evolution of warning coloration? *Oikos* **45**: 31–36.

Guilford, T. 1988. The Evolution of Conspicuous Coloration. *Am. Nat.* **131**: S7–S21.

Haig, D. 1996. Gestational drive and the green-bearded placenta. *Proc. Natl. Acad. Sci.* **93**: 6547–6551.

Haig, D. 2013. Imprinted green beards: a little less than kin and more than kind. *Biol. Lett.* **9**: 20130199.

Haig, D. 1997. The social gene. In: *Behavioural Ecology: an evolutionary approach* (J. R. Krebs & N. B. Davies, eds), pp. 284–304. Blackwell Science Ltd.

Hamilton, W.D. 1972. Altruism and Related Phenomena, Mainly in Social Insects. *Annu. Rev. Ecol. Syst.* **3**: 193–232.

Hamilton, W.D. 1975. Innate Social Aptitudes of Man: an approach from evolutionary genetics. In: *ASA Studies 4: Biosocial Anthropology* (R. Fox, ed), pp. 133–153. Malaby Press.

Hamilton, W.D. 1964a. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**: 1–16.

Hamilton, W.D. 1964b. The genetical evolution of social behaviour II. *J. Theor. Biol.* **7**: 17–52.

Hamilton, W.D. 2001. *Narrow Roads of Gene Land: Volume 2*. Oxford University Press.

Heller, J., Zhao, J., Rosenfield, G., Kowbel, D.J., Gladieux, P. & Glass, N.L. 2016. Characterization of Greenbeard Genes Involved in Long-Distance Kind Discrimination in a Microbial Eukaryote. *PLoS Biol.* **14**: e1002431.

Hirose, S., Benabentos, R., Ho, H.-I., Kuspa, A. & Shaulsky, G. 2011. Self-Recognition in Social Amoebae Is Mediated by Allelic Pairs of Tiger Genes. *Science (80-.).* 333: 467–470.

Huang, Y.C. & Wang, J. 2014. Did the fire ant supergene evolve selfishly or socially? *BioEssays* **36**: 200–208.

Hurst, G.D.D. & McVean, G.A.T. 1998. Selfish genes in a social insect. *Trends Ecol. Evol.* **13**: 434–435.

Jansen, V.A.A. & van Baalen, M. 2006. Altruism through beard chromodynamics. *Nature* **440**: 663–666.

Karadge, U.B., Gosto, M. & Nicotra, M.L. 2015. Allorecognition Proteins in an Invertebrate Exhibit Homophilic Interactions. *Curr. Biol.* **25**: 2845–2850.

Keller, L. 2002. Marsupial sperm pairing: A case of "sticky" green beards? *Trends Ecol. Evol.* **17**: 112–113.

Keller, L. & Ross, K.G. 1998. Selfish genes: a green beard in the red fire ant. *Nature* **394**: 573–575.

Krupp, D.B. & Taylor, P.D. 2015. Social evolution in the shadow of asymmetrical relatedness. *Proc. R. Soc. B Biol. Sci.* **282**: 20150142.

Linksvayer, T.A., Busch, J.W. & Smith, C.R. 2013. Social supergenes of superorganisms: do supergenes play important roles in social evolution? *Bioessays* **35**: 683–689.

Marshall, J.A.R. 2015. *Social Evolution and Inclusive Fitness Theory*. Princeton University Press.

Nicotra, M.L., Powell, A.E., Rosengarten, R.D., Moreno, M., Grimwood, J., Lakkis, F.G., et al. 2009. A Hypervariable Invertebrate Allodeterminant. *Curr. Biol.* **19**: 583–589.

Nydam, M.L., Netuschil, N., Sanders, E., Langenbacher, A., Lewis, D.D., Taketa, D.A., et al. 2013. The Candidate Histocompatibility Locus of a Basal Chordate Encodes Two Highly Polymorphic Proteins. *PLoS One* **8**: e65980.

Pathak, D.T., Wei, X., Dey, A. & Wall, D. 2013. Molecular Recognition by a Polymorphic Cell Surface Receptor Governs Cooperative Behaviors in Bacteria. *PLoS Genet.* **9**: e1003891.

Pepper, J.W. & Smuts, B.B. 2002. A mechanism for the evolution of altruism among nonkin: Positive assortment through environmental feedback. *Am. Nat.* **160**: 205–213.

Pollak, S., Omer-Bendori, S., Even-Tov, E., Lipsman, V., Bareia, T., Ben-Zion, I., et al. 2016. Facultative cheating supports the coexistence of diverse quorum-sensing alleles. *Proc. Natl. Acad. Sci.* **113**: 2152–2157.

Pracana, R., Levantis, I., Martínez-Ruiz, C., Stolle, E., Priyam, A. & Wurm, Y. 2017. Fire ant social chromosomes: Differences in number, sequence and expression of odorant binding proteins. *Evol. Lett.* **1**: 199–210.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. 2014. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**: 2728–2732.

Queller, D.C. 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. *Proc. Natl. Acad. Sci.* **108**: 10792–10799.

Queller, D.C. 1984. Kin selection and frequency dependence: a game theoretic approach. *Biol. J. Linn. Soc.* **23**: 133–143.

Queller, D.C. 1992. Quantitative Genetics, Inclusive Fitness, and Group Selection. *Am. Nat.* **139**: 540–558.

Queller, D.C., Ponte, E., Bozzaro, S. & Strassmann, J.E. 2003. Single-Gene Greenbeard Effects in the Social Amoeba Dictyostelium discoideum. *Science (80-.).* **299**: 105–106.

Quickfall, C.G. & Marshall, J.A.R. 2017. The evolution of mutualism with modifiers. *Ecol. Evol.* **7**: 6114–6118.

Ratnieks, F.L.W., Foster, K.R. & Wenseleers, T. 2006. Conflict Resolution in Insect Societies. *Annu. Rev. Entomol.* **51**: 581–608.

Richard, F., Glass, N.L. & Pringle, A. 2012. Cooperation among germinating spores facilitates the growth of the fungus, Neurospora crassa. *Biol. Lett.* **8**: 419–422.

Ridley, M. & Grafen, A. 1981. Are green beard genes outlaws? *Anim. Behav.* **29**: 954-955.

Rothstein, S.I. & Barash, D.P. 1983. Gene conflicts and the concepts of outlaw and sheriff alleles. *J. Soc. Biol. Syst.* 6: 367–379.

Scofield, V.L., Schlumpberger, J.M., West, L.A. & Weissman, I.L. 1982. Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* **295**: 499–502.

Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Guadagnini, S., Yan, C., et al. 2008. FLO1 Is a Variable Green Beard Gene that Drives Biofilm-like Cooperation in Budding Yeast. *Cell* **135**: 726–737.

Springer, S.A., Crespi, B.J. & Swanson, W.J. 2011. Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture. *Mol. Ecol.* **20**: 2240–2257.

Thompson, M.J. & Jiggins, C.D. 2014. Supergenes and their role in evolution. *Heredity (Edinb).* **113**: 1–8.

Trible, W. & Ross, K.G. 2015. Chemical communication of queen supergene status in an ant. *J. Evol. Biol.* **29**: 502–513.

Unterweger, D. & Griffin, A.S. 2016. Nice or nasty: Protein translocation between bacteria and the different forms of response. *Proc. Natl. Acad. Sci.* **113**: 8559–8561.

Vassallo, C.N., Cao, P., Conklin, A., Finkelstein, H., Hayes, C.S. & Wall, D. 2017. Infectious polymorphic toxins delivered by outer membrane exchange discriminate kin in myxobacteria. *Elife* **6**: e29397.

Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **37**: 986–990.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., et al. 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**: 664–8.

West, S.A. & Gardner, A. 2013. Adaptation and inclusive fitness. *Curr. Biol.* **23**: R577-84.

West, S.A., Griffin, A.S. & Gardner, A. 2007. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: R661–R672.

West, S.A., Griffin, A.S., Gardner, A. & Diggle, S.P. 2006. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4**: 597–607.

Wielgoss, S., Fiegna, F., Rendueles, O., Yu, Y.T.N. & Velicer, G.J. 2018. Kin discrimination and outer membrane exchange in Myxococcus xanthus: A comparative analysis among natural isolates. *Mol. Ecol.* **27**: 3146–3158.

Zhang, H. & Chen, S. 2016. Tag-mediated cooperation with non-deterministic genotype-phenotype mapping. *Europhys. Lett.* **113**: 28008.

Comment 4: Population genetics of greenbeard concepts

Following on from the general absence of data about the mechanisms of conditional cooperation, the preceding paper used current theory and the latest experimental findings to evaluate and hone the empirical utility of the greenbeard concept, setting up new questions for future research. I leave those questions for the future, but here I build on some of the verbal arguments within Paper 4 with simple mathematical models to elaborate on the central narrative and its implications for the evolution of conditional cooperation.

Paper 4 argued that the Dawkins (1976) presentation of the greenbeard thought experiment has had an enormous impact on how the fundamental principle within the greenbeard concept is understood. Much of the subsequent literature has analysed the greenbeard concept through the framework of indirect fitness accounting (*e.g.* West *et al.*, 2007a), which emphasises how greenbeards are susceptible to falsebeards (Dawkins, 1982; Gardner & West, 2010). Indeed, this vulnerability is the most-often cited reason why greenbeards would not be anticipated in nature, and to that extent is wholly misleading given that greenbeards do appear to be plausibly found in nature. When Hamilton (1964b) first envisaged the greenbeard concept, he did not use the framework of inclusive fitness to make this case, but instead verbally presented a different kind of model that is much more simplistic. Hamilton's description was very similar to Dawkins' (1976), but had an additional part that discussed a cost-free concept that Dawkins dropped in *The Selfish Gene*. The cost-free concept is closer to the 'master regulator' concept that we emphasised in Paper 4, highlighting that a greenbeard need not directly cause the social behaviour in the manner that Dawkins (1976) imagined. An important implication of this concept is how greenbeard genes can run into intragenomic conflict with kin-selected 'modifier' genes (Ridley & Grafen, 1981; Biernaskie *et al.*, 2011). Lastly, when Haig (1996) reinvented the greenbeard concept for empirical research by suggesting how it could be achieved by molecular mechanisms, he presented the greenbeard thought experiment through the lens of Dawkins (1976). However, Haig (1996)

also verbally presented a variant concept within another framework to emphasise the link between greenbeards and meiotic drive, which emphasises how greenbeards can be selected despite harming individual fitness. Here, I describe and contrast these three variant presentations of the greenbeard concept in Hamilton (1964b), Dawkins (1976) and Haig (1996) to highlight their different implications for understanding conditional cooperation.

When Hamilton (1964b) first proposed the greenbeard concept (within a paper outlining the logic of inclusive fitness), the greenbeard gene was conceived of as influencing the recipient of a social behaviour rather than directly causing the social behaviour.

"*That genes could cause the perception of the presence of like genes in other individuals may sound improbable; at simplest we need to postulate something like a supergene affecting (a) some perceptible feature of the organism, (b) the perception of that feature, and (c) the social response consequent upon what was received. … If some sort of attraction between likes for purposes of co-operation can occur the limits to the evolution of altruism expressed by our first principle would be very greatly extended, although it should still never happen that one individual would value another more highly than itself, fitness for fitness. And if an individual can be attracted towards likes when it has positive effects – benefits – to dispense, it can presumably be attracted the other way, towards unlikes, when it has negative effects to dispense (i.e. when circumstances arise which demand combat, suggest robbery, and so on).*" (Hamilton 1964b, p.25-26)

After presenting the three requirements, Hamilton suggests that if an individual has cooperative benefits to give to a social partner then a gene governing the recipients of this behaviour might do better to give those benefits to another individual that shares that gene. In this second part of the idea's exposition, the greenbeard gene only determines the probability that both the actor and the recipient share the greenbeard gene, which is their coefficient of relatedness – making the social behaviour itself cost-free. The greenbeard gene does not need to cause the social behaviour that they direct toward some recipients rather than others (which

is in-keeping with the master regulator concept of Paper 4, but we do not restrict our attention exclusively to refer to conditional social behaviours that are cost-free).

To examine this verbal model in greater detail, we can consider a pair of alleles in a simple haploid model. For simplicity, we only discuss the case where individuals have positive effects ($b > 0$) to dispense to others because it is simpler than the case where individuals have negative effects to dispense (which requires negative relatedness) because consideration of the first case is enough to demonstrate my point. The background relatedness between alleles is zero because we assume no population structure, but the greenbeard allele increases the relatedness to r (which is interpreted geometrically as the probability of sharing the greenbeard allele over and above a random expectation). In this way, the greenbeard carriers (of frequency p) give out their benefits to a non-random subset of individuals ($r + (1 - r)p$) whilst nonbeard carriers give out their benefits indiscriminately ($1 - p$), so the fitness of a greenbeard allele against a nonbeard allele is:

$$\omega_{GB} = 1 + \big((r + (1 - r)p) + (1 - p)\big)b \tag{1.1}$$

$$\omega_{NB} = 1 + \big((1 - r)(1 - p) + (1 - p)\big)b \tag{1.2}$$

Thus $\omega_{GB} > \omega_{NB}$ whenever $r > 0$. Within this framework, it is not clear what it would mean for an allele to be a falsebeard. An allele could reduce r (pushing it closer to zero, *i.e.* indiscriminately giving out benefits) but then it is hard to intuit how this allele could disrupt the greenbeard-driven assortment of benefits to itself (by pushing r to be less than zero). Therefore, at worst, a falsebeard would be neutral to a greenbeard in assorting to receive benefits from greenbeard actors whilst indiscriminately giving benefits to all alleles.

Hamilton's (1964b) cost-free presentation of the greenbeard thought experiment implies a 'master regulator' like role for greenbeards. Instead of the greenbeard producing

some beneficial behavioural effect for those that share the gene, a greenbeard simply directs a beneficial behaviour (that it has no role in producing) to those that share the gene. Consequently, we might imagine a gene to be the 'decision-maker' rather than the 'decision-taker' (*cf.* Bachrach & Baratz, 1962), in soliciting the behavioural effect that other genes functionally work to carry out despite potentially gaining no benefit from doing so (if they are unlinked). Those other genes may be pleiotropically constrained as part of some general behavioural mechanism and so be unable to prevent the greenbeard's corruption, but it seems more likely that these genes could evolve as modifiers to become resistant to the greenbeard. A gene that is linked to the greenbeard locus would receive a diluted benefit in correspondence with its level of linkage, but an unlinked modifier would make no gains from the greenbeard's regulatory intervention. Because, in equation (1), the behaviour is cost-free, an unlinked modifier that fully suppresses the greenbeard (*i.e.* ignores its solicitation) would not be favoured by natural selection. However, if a kin-selected decision-making gene were in competition with the greenbeard over the regulation of this behavioural effect, the unlinked genes in the genome could be selected to fully suppress the greenbeard regulator and permit the kin-selected regulator. In this way, even if a greenbeard is cost-free, it could lead to intragenomic conflict between the greenbeard locus (and other linked loci) and the rest of the genome.

Dawkins (1976) presentation of Hamilton's (1964b) original thought experiment, which gave the concept its name, focused on the first part of the thought experiment and placed less emphasis on the verbal model that was presented in the second part (see above for comparison):

"*It is theoretically possible that a gene could arise which conferred an externally visible 'label', say a pale skin, or a green beard, or anything conspicuous, and also a tendency to be specially nice to bearers of that conspicuous label. It is possible, but not particularly likely. … It is not very probable that one and the same gene would produce both the right label and the right sort of altruism. Nevertheless, what may be called the Green Beard Altruism Effect*

is a theoretical possibility. An arbitrary label like a green beard is just one way in which a gene might 'recognize' copies of itself in other individuals." (Dawkins 1976, p.89)

Following on from this presentation, greenbeards can be understood as genes for altruism, which is clearly different from Hamilton's verbal model. Altruism can be conceptually analysed in terms of their inclusive fitness accounting, which immediately emphasises the problem of the vulnerability of greenbeards to 'falsebeards' that cheat the greenbeard by presenting the right signal but not responding with altruism to others that bear the right signal. Thus a greenbeard receives a benefit of altruism ($b > 0$) when it interacts with itself (at its population-wide frequency p), but always pays the cost of altruism ($c > 0$), whilst a falsebeard only ever receives the benefit of altruism ($b > 0$) when it interacts with the greenbeard (at frequency p). For the sake of parity with the simple master regulator model presented above, we can consider some background level of relatedness between social partners (r), so that greenbeard and falsebeard fitness is:

$$\omega_{GB} = 1 + (r + (1 - r)p)b - c \tag{2.1}$$

$$\omega_{FB} = 1 + (1 - r)pb \tag{2.2}$$

Thus $\omega_{GB} > \omega_{FB}$ whenever $rb - c > 0$. Further, if a greenbeard could accurately detect itself, then it would be able to invade a nonbeard ($\omega_{GB} = 1 + (r + (1 - r)p)(b - c)$; $\omega_{NB} = 1$), and so as long as $r > 0$ then $\omega_{GB} > \omega_{NB}$ reduces to the most favourable expression $b - c > 0$.

Haig (1996) presents another verbal (but mathematically oriented) model to demonstrate the similarities between greenbeard genes and meiotic drive genes:

"*Suppose that a mother has limited resources available for reproduction and produces offspring one at a time. The less she invests in each offspring, the more offspring she produces, but the lower the probability of survival for each individual offspring. The amount of investment per offspring that maximizes maternal fitness can be represented by m*, the quantity that optimizes the trade-off between offspring number and offspring quality … . Consider a single* Dd *female in a population of* dd *individuals. If meiosis is fair, each of this female's offspring has an equal chance of being* Dd *or* dd*. All other mothers produce an unbroken sequence of* dd *offspring, each of whom receives m* (by assumption). Therefore, if* D *causes offspring to receive m*+δ, the* Dd *mother will leave fewer surviving offspring than a* dd *mother, even though the offspring who receive the extra amount have enhanced survival. … The analysis also clarifies the conceptual similarity between models of meiotic drive and models of parent-offspring conflict. In the former, genetic agents are able to invade a population because they distort the process of meiosis to gain access to more than 50% of successful gametes. In the latter, genes that cause increased offspring demands are able to invade a population because they distort the process of parental care to gain more than their fair share of resources.*" P.6547

This description suggests that greenbeards can be selected despite imposing a cost on the genes in the rest of the genome (in explicitly constructing the greenbeard gene as a selfish genetic element). This statement can be made even more explicit by considering a simpler meiotic drive model (without a maternal effect). In this setup, a meiotic drive-like greenbeard (via obligate harm) reaps a transmission advantage against a nonbeard when in a heterozygote but suffers a selective disadvantage when in a homozygote (see Paper 5). In terms of inclusive fitness, a greenbeard always carries a fixed advantage to the actor (*i.e.* a negative cost; $c < 0$) but reaps a conditional disadvantage (*i.e.* a negative benefit; $b < 0$) when in a homozygote (which otherwise goes to the nonbeard). The likelihood of a heterozygote forming is governed by relatedness (r) as before, but here this parameter might better understood as reflecting the degree of inbreeding. In this way, the fitness of a greenbeard and nonbeard allele is:

$$\omega_{GB} = 1 + (r + (1 - r)p)b - c \quad (3.1)$$

$$\omega_{NB} = 1 + (1 - r)pb \quad (3.2)$$

Because meiotic drive relies on an 'unfair' advantage to the greenbeard in the heterozygote, relatedness has the opposite effect on whether or not the greenbeard or nonbeard allele goes to fixation. The condition for $\omega_{GB} > \omega_{NB}$ leads to Hamilton's rule, $rb - c > 0$, similar to the previous scenario but in this scenario we are supposing $b < 0$ and $c < 0$. Further, Equation (3.2) may look similar to Equation (2.2,) but Equation (3.2) is for a nonbeard allele not a falsebeard allele; consequently, this might imply that there is no conditionality, or alternatively that (like Equation 1) there is no real distinction between a falsebeard and nonbeard allele – as any allele can act like a falsebeard. But importantly, for appropriate values when $b < 0$ and $c < 0$, the greenbeard can go to fixation despite imposing a load on the population (of $b - c$). In this way, a gene can be selected despite leading to maladaptation in decreasing individuals' fitness to those that carry the greenbeard.

These three models of greenbeard genes focus on different parts of the fundamental idea's conceptualisation. Hamilton's description draws attention to greenbeards as genes that change the relatedness between social partners. Dawkins' description draws attention to greenbeards as genes for social behaviour (especially altruism). Haig's description draws attention to greenbeards as selfish genetic elements. The definition of a greenbeard gene in Paper 4 drew on each of these descriptions. However, current empirical research has primarily focused on finding evidence for Dawkins' conceptualisation, despite the details of this evidence often arguably providing better support for Hamilton's conceptualisation (see Table 2 for Hamiltonian 'informative' vs Dawkinsian 'causal' in Paper 4). But the broader point of Paper 4 is to direct attention to a more fundamental concept of what is greenbeard is. Much of the confusion surrounding greenbeard genes (*e.g.* whether greenbeard genes are outlaws) is

likely to have its origins in the way in which the concept can be constructed in different frameworks (as these three cases demonstrate), which each bring different evolutionary questions to the fore. However, we should be very clear that each of these perspectives correctly describes a different aspect of the greenbeard concept, and therefore there is a great need to establish how general any claims about particular greenbeard genes are within different (but supposedly general) modelling frameworks. Each of these concepts has been framed to analyse the balance of selection on a gene within the inclusive fitness partitioning of direct fitness effects to the actor (c) and indirect fitness effects via the recipient (b). In this framework, the master regulator greenbeard has a negligible direct fitness effect ($c \to 0$), the indirect fitness greenbeard is for altruistic social behaviour ($c < 0$ and $b > 0$; though spite has also been considered elsewhere) and the meiotic drive greenbeard is for selfish social behaviour ($c > 0$ and $b < 0$; though mutual benefit has also been considered elsewhere).

The evolutionary understanding of how conditional cooperation plays out in nature is still in its early days. As Paper 4 emphasised, all of the examples that are currently suggested to be greenbeards could turn out to be kin-selected because the critical 'naturalised' experiments to rule out kin selection are absent. Of the examples given, the *tgr* gene in *Dictyostelium discoideum* is perhaps the most likely to be a greenbeard, given that Tgr protein binding strength correlates with *tgr* genetic distance but not genome-wide genetic distance (Gruenheit *et al.*, 2017). However, I would suggest that the presence of polymorphism at the locus that encodes a signal and a specific complementary receptor governing a social effect presents strong evidence that a gene is a greenbeard because I do not see how this mechanism could detect genome-wide relatedness (or when the signal and the receptor are the same self-binding protein; see also Haig 1996). The same could not be said when the locus does not encode the signal (*e.g.* relying upon a pleiotropically-constrained cue from some unlinked locus) because this introduces some uncertainty about the possession of the gene in in question. Further, for many systems, the greenbeard properties may simply be uninteresting (*e.g.* there is no intragenomic conflict) making there be no advantage to applying this concept.

To-date, the evidence of the genetic/phenotypic mechanism of conditional cooperation is best exemplified in many of these candidate greenbeard systems, and so it seems likely, given the number of different examples with signal-receptor polymorphism that are discussed in Paper 4, that greenbeards could have an important part to play in the evolutionary explanation of conditional cooperation.

Taking inspiration from the logic of greenbeards, I next want to move to consider how these ideas about conditional cooperation can be applied to improve the design of gene drives. All of the different conceptions of greenbeard genes are presented in a way that suggests that that greenbeards are likely to be short-lived (Hamilton, 1964b; Dawkins, 1976; Haig, 1996) and yet, if a greenbeard and a kin-selected allele were in direct competition for control of the recipients of a social behaviour, presumably the greenbeard allele might be expected to outcompete the kin-selected allele because (paradigmatically) the greenbeard allele can ascertain which individuals share that particular allele rather than the kin-selected allele that can only ascertain some probability of sharing the allele due to common ancestry (*i.e.* the greenbeard has a higher relatedness coefficient). In this way, a greenbeard is a good allelic competitor, but is thought to be subject to modifiers at other loci that suppress its activity (Ridley & Grafen, 1981; Biernaskie *et al.*, 2011). Is it possible to have a gene that is both a proficient allelic competitor and immune to suppressing modifiers? I discuss this in the applied context of gene drive technologies.

Paper 5

This declaration concerns the article entitled:			
Designing gene drives to evade rapidly-evolving resistance			
Publication status (tick one) **Draft manuscript** ☐ **Submitted** ✓ **In review** ☐ **Accepted** ☐ **Published** ☐			
Publication details (reference)	**Madgwick, P. G.**, and J. B. Wolf. n.d.. Designing gene drives to evade rapidly-evolving resistance.		
Copyright status (tick the appropriate statement) I hold the copyright for this material ✓ Copyright is retained by the publisher, but I have been given permission to replicate the material here ☐			
Candidate's contribution to the paper (provide details, and also indicate as a percentage)	Formulation of ideas: 80%: PGM significantly contributed to the concept, literature review and narrative opinion. Design of methodology: N/A Experimental work: N/A Presentation of data in journal format: 80%: PGM predominantly executed the drafting, revising and editing of the manuscript.		
Statement from Candidate	This paper reports on original research I conducted during the period of my Higher Degree by Research candidature.		
Signed	PGMadgwick	**Date**	07/11/2019

Data access statement: N/A

Designing gene drives to evade rapidly-evolving resistance

Authors: Philip G. Madgwick[1*], and Jason B. Wolf[1]

Affiliations: [1] Milner Centre for Evolution and Department of Biology & Biochemistry, University of Bath, Bath, BA2 7AY, UK

[*] Correspondence to: pgm29@bath.ac.uk

Abstract

Gene drive technologies have attracted attention for providing a means of controlling disease vectors and invasive pest populations (Burt, 2003, 2014; Champer *et al.*, 2016; Godfray *et al.*, 2017; Marshall & Akbari, 2018; McFarlane *et al.*, 2018). Gene drives can be used to suppress or exterminate natural populations using a selfish genetic element that propagates through a population despite incurring a cost to individual fitness. A major hurdle for proposed forms of gene drives is the rapid evolution of resistance (Champer *et al.*, 2017, 2018; Hammond *et al.*, 2017; Noble *et al.*, 2017, 2018; Unckless *et al.*, 2017), which can limit their potential utility. Here, we propose how gene drives can be designed to reduce the risk of the rapid evolution of resistance (whilst still being controllable).

Keywords: gamete-killer, meiotic drive, selfish genetic elements, gene conflict, population control

Main-text

Population genetic models suggest that gene drives are expected to rapidly spread through the population (Sandler & Novitski, 1957; Leigh, 1971; Charlesworth & Hartl, 1978;

Deredec *et al.*, 2008), which is supported by natural examples (Palopoli & Wu, 1996; Hornett *et al.*, 2009; Helleu *et al.*, 2016) and laboratory experiments (Lyttle, 1977; Price *et al.*, 2010; Pinzone & Dyer, 2013). The spread of a drive allele creates a strong selective pressure for the rapid evolution of resistance to its effects (Champer *et al.*, 2017, 2018; Hammond *et al.*, 2017; Unckless *et al.*, 2017; Noble *et al.*, 2018). For some types of gene drive that rely upon 'over-replication' (McLaughlin & Malik, 2017), including the (CRSIPR) homing endonucleases that have received the most attention (Burt, 2003, 2014; Champer *et al.*, 2016; Godfray *et al.*, 2017; Marshall & Akbari, 2018; McFarlane *et al.*, 2018), errors in the process whereby they replicate between sites within the genome can readily lead to the creation of a resistant allele – which has been experimentally shown to occur on very short timescales (Hammond *et al.*, 2017; Noble *et al.*, 2017; Unckless *et al.*, 2017; Champer *et al.*, 2018). Obviously, for a gene drive to be successful, it is essential to minimize the risks of resistance rapidly evolving, and there are means by which over-replication gene drives can do this (Champer *et al.*, 2017, 2018; Hammond *et al.*, 2017; Unckless *et al.*, 2017; Noble *et al.*, 2018). However, no matter how much these techniques reduce the effects of replication error, resistance is still very likely to evolve for over-replication gene drives in the shorter-term because they create a selective advantage for a modifier at another locus elsewhere in the genome that suppresses their activity (Prout *et al.*, 1973; Hartl, 1975; Thomson & Feldman, 1976). In this way, recent advances in the theory of gene drives have tended to prioritise maximizing the possible extent of population suppression (Burt, 2003, 2014; Champer *et al.*, 2016; Godfray *et al.*, 2017; Marshall & Akbari, 2018; McFarlane *et al.*, 2018), rather than minimising the risk of resistance – both of which are necessary for an effective gene drive (Bull & Malik, 2017).

Although over-replication gene drives have received the greatest attention (Burt, 2003, 2014; Champer *et al.*, 2016; Godfray *et al.*, 2017; Marshall & Akbari, 2018; McFarlane *et al.*, 2018), the potential for rapidly evolving resistance in these systems is highly undesirable, and so it is worthwhile considering whether any alternatives would be more effective at minimizing the risk of resistance. Other types of selfish genetic element that rely

upon transmission distortion (McLaughlin & Malik, 2017) could also be used for gene drives, such as a zygote- or gamete-killer. Transmission distorters replicate alongside the rest of the genome, but they bias the process of gamete production to ensure that offspring have an elevated probability of possessing the transmission distorter, which often comes at a cost of reducing their carrier's fertility. In these systems, the most common source of resistance in the shorter-term is not likely to be the evolution of a resistant allele at that locus (because there is no error-prone over-replication process) but, rather, the evolution of modifier alleles at other loci that suppress the transmission distorter and restore fairness to sexual reproduction (because modifiers are simply a larger mutational target). Of the selfish genetic elements for transmission distortion, the zygote-killer *Medea* (Beeman *et al.*, 1992; Wade & Beeman, 1994; Buchman *et al.*, 2018) has received the most attention for gene drive because it is relatively easy to genetically engineer. However, like homing endonucleases, *Medea* is just as susceptible to suppression by modifiers at other loci (Prout *et al.*, 1973; Hartl, 1975; Thomson & Feldman, 1976). Therefore, moving beyond the 'usual suspects', it is interesting to note that by far the most numerous form of selfish genetic element that have been detected in natural populations are gamete-killers (Burt & Trivers, 2008; Lindholm *et al.*, 2016), which may suggest that they have some ability to evade the evolution of resistance and persist, at the very least, on shorter timescales. Here, we argue that some mechanisms of gamete-killing are more likely to evade the rapid evolution of resistance than others and, for this reason, we put forward one mechanism of gamete-killing that may potentially generate a more effective gene drive for controlling natural populations.

Let us imagine a gamete-killing gene drive that uses a toxin-antitoxin system to generally refer to both killer-target or poison-antidote systems (Bravo Núñez *et al.*, 2018), which could plausibly be genetically engineered from current methods (*e.g.* Chen *et al.*, 2007). We can consider the three types of killer depending on when the toxin is expressed and takes effect during meiosis (Haig & Grafen, 1991), which determines how cells in different haploid and diploid combinations interact (see Figure 1). Logically, for any invadable killer gene

drive, the toxin must always take effect by killing a haploid sister or daughter cell because otherwise killing a diploid parental cell would render a heterozygote infertile (and the killer could not invade). Consequently, the three types of killers reflect the only viable routes for a gamete-killing gene drive. First, the toxin can be expressed before meiosis in a diploid parental cell (or, indeed, some other somatic gonadal cells), but have its effect on the haploid sister or daughter cells. Second, the toxin can be expressed by and have an effect on the haploid sister cells after the first reductional division of meiosis and before the second equational division of meiosis. Third, the toxin can be expressed by the haploid daughter cells (*i.e.* the gametes) after the second equational division at the end of meiosis, having an effect on either the haploid sister or daughter cells. The difference between the second and third types is in the way in which the cells expressing the toxin interact, with sister cells interacting as a dyad rather than daughters cells interacting as a tetrad. We refer to these three activity profiles of expression and effect as pre-meiotic, meso-meiotic, and post-meiotic killing respectively.

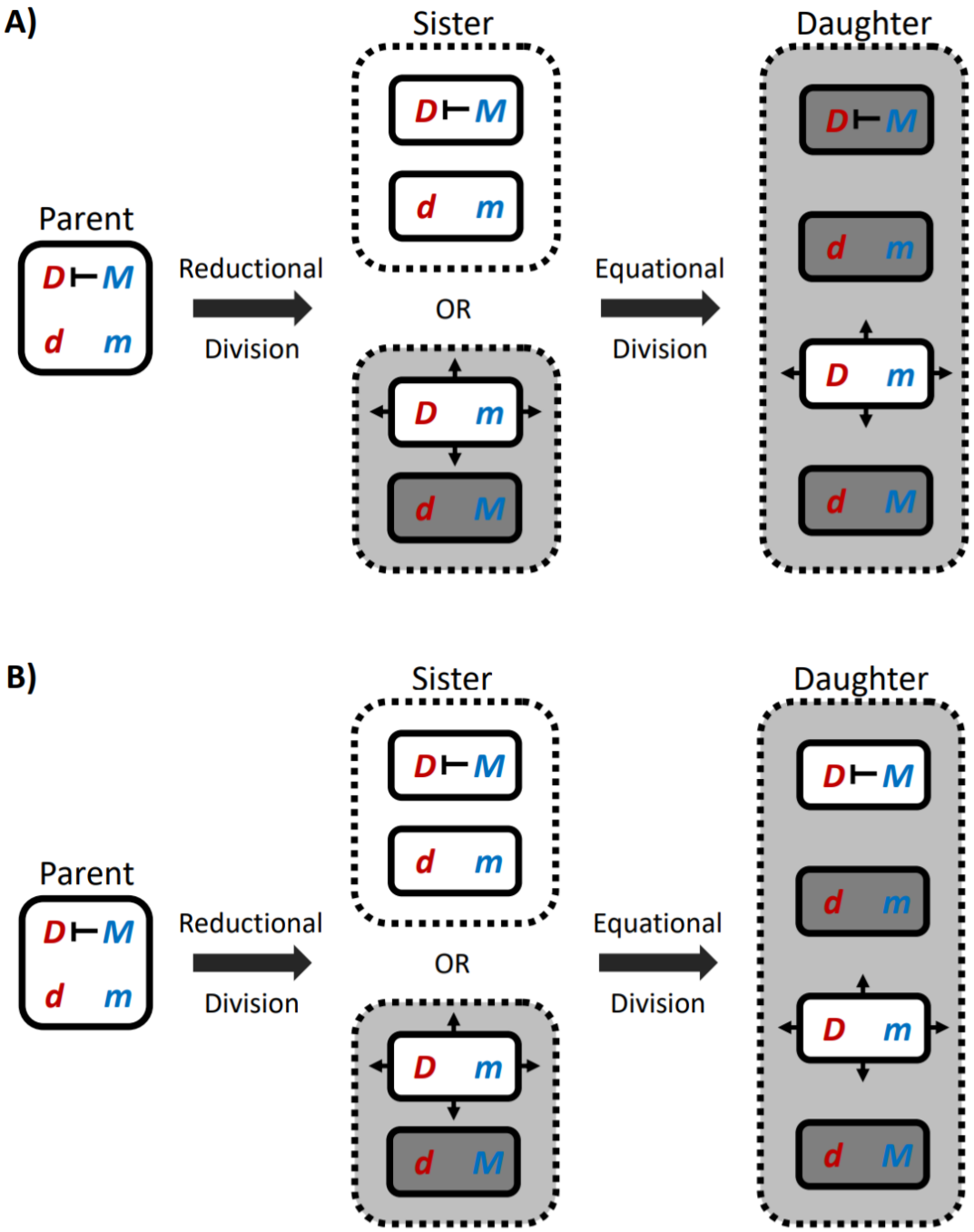


Figure 1. A simplified diagram showing the potential for different types of game-killing gene drive throughout the steps of the meiotic divisions in a double heterozygote (patterns in other genotypes are illustrated in SI Appendix, Figure S3). During meiosis, a diploid parental cell undergoes a first reductional division into two haploid sister cells followed by a second (mitosis-like) equational division into four haploid daughter cells. Different cells are presented with solid outlines, and different interactions between cells are presented with dashed outlines

(except for the possible interaction between the parent cell and the sister cells it produced). Alleles at the drive locus are presented in red (D = drive and d = nondrive) and alleles at the modifier locus are presented in blue (M = modifier and m = nonmodifier). A drive allele for a toxin-antitoxin pair can distort its transmission to the next generation by producing a toxin (indicated by out-facing arrows from the producer cell) that leads to the presence of the toxin within that given interaction (indicated by the light-grey shading in dashed outlined box). The presence of the toxin kills gametes that do not also produce the antixotin, with cell death indicated by the dark-grey shading in solid outlined boxes). The modifier allele (M) only has an effect on the drive (D) allele if it is within the same cell, which is indicated by the turnstile symbol ⊢ that indicates that the drive allele has blocked some function of the drive allele (with different scenarios illustrated in A and B). As such, in the double heterozygote scenario presented here, there is no potential for pre-meiotic killing, where the diploid parent cell produces the toxin that kills haploid sister or daughter cells, because the modifier blocks toxin production. However, there is potential for meso-meiotic killing between sister cells (of which there are two potential genotypic combinations represented by different interactions in dashed boxes) and post-meiotic killing between daughter cells (ignoring the potential for daughter cells to kill sister cells in this simplified diagram). **A)** The case of a fully suppressing modifier that blocks toxin and antitoxin production by the drive allele, rendering it susceptible to the toxin produced by other cells. **B)** The scenario of a partially suppressing modifier blocks that toxin production only, and so remains resistant to the toxin produced by other cells.

As a one-locus system, all three forms of killing are phenotypically equivalent. This basic scenario can be modelled using a simple framework (following Leigh, 1971) where a drive allele (D, with frequency f_D) is competing against a nondrive allele (d, with frequency f_d). Whenever an individual carries the drive allele, their gametes suffer a selective disadvantage ($-s$) from being poisoned by the toxin in spite of the antitoxin (where this cost is scaled to vary between 0 and 1), which provides a component of hard selection against the drive allele. In this way, the toxicity cost ($-s$) determines the magnitude of the fitness load (which is expected to be manifested as reduced fecundity, and hence represents the utility of the drive allele for population suppression) when the drive allele is at fixation in the population, reducing individual fitness from 1 to $1-s$ (Figure 2). The drive allele also imposes a transmission disadvantage on the nondrive allele ($-e$) that arises from killing of nondrive carrying gametes by the toxin, which gains the drive allele a transmission advantage ($+e$), and hence the drive allele can potentially promote its own transmission despite the toxicity cost through this component of soft selection. The transmission advantage, which only arises in heterozygotes (because it requires an interaction between the drive and non-drive alleles), is scaled from a minimum of 0 for a fair meiosis and a maximum of 1 for the case where only drive allele carrying gametes survive. In heterozygotes, the toxicity cost ($-s$) and transmission effect ($\pm e$) are multiplicative because the former determines the number of viable gametes produced (hence representing hard selection) whilst the latter determines the proportion of those gametes that contain the drive allele (hence representing soft selection; Table 1). Therefore, the overall fitness of the drive allele is the frequency-weighted average of its fitness when homozygous (where it pays the toxicity cost) and its fitness when heterozygous (where it pays the cost, but also gains the transmission advantage), with corresponding implications for the nondrive allele:

$$\omega_D = 1 - s + f_d e(1 - s) \quad (1a)$$

$$\omega_d = 1 - f_D s - f_D e(1 - s) \quad (1b)$$

Therefore, the drive allele is favoured when:

$$f_D > \frac{1}{s}\left(1 - (1 + e)(1 - s)\right) \quad (2)$$

Logically, recapitulating a classic result (Leigh, 1971), the drive allele is always able to invade whenever the product of the toxicity cost and transmission advantage is greater than one (*i.e.* $(1 + e)(1 - s) > 1$; or $e > s/(1 - s)$; see SI Appendix, Figure 1S). Consequently, the drive allele will be more likely to be favoured (and selection favouring the drive allele would be stronger) the smaller the selective disadvantage ($-s$) and the larger the transmission effect (e). These conditions highlight the challenge of population control via a gamete-killing drive allele. As the selective disadvantage ($-s$) is expected to manifest itself as a fecundity cost, there is a trade-off between the capacity for the drive allele to invade and to suppress the population: the lower the fecundity cost the less restrictive the invasion criteria, but also the lower the impact on mean fitness as the drive allele goes to fixation (where population mean fitness goes from 1 in the starting population to $1 - s$ when the drive allele is at fixation). If there is a negligible fecundity cost, the drive allele can easily invade and spread to fixation against the nondrive allele, but this accomplishes nothing with respect to population control. To provide the most effective population suppression, the fecundity cost has to be as large as possible, but s has to be less than ½ for the drive allele to invade. Consequently, the drive allele cannot cause a population cannot to lose more than half of its fecundity, which limits the scope of population suppression through this approach.

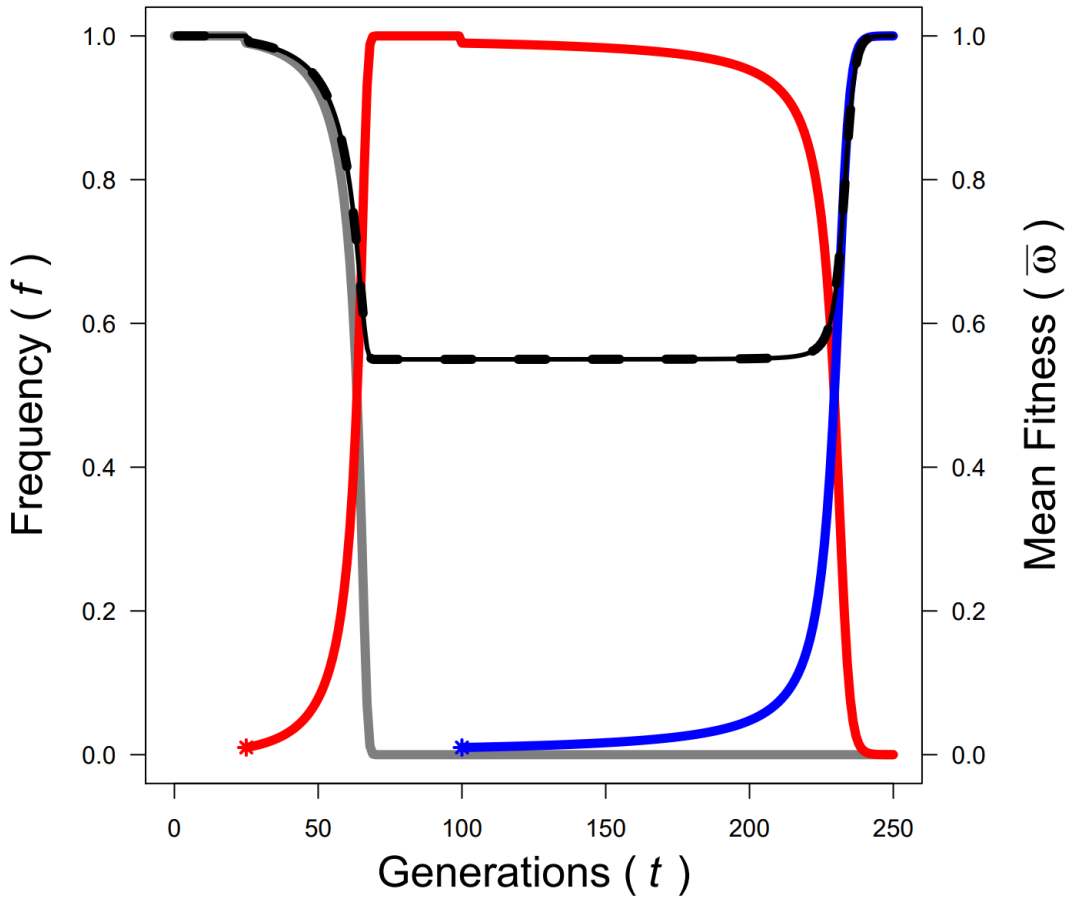


Figure 2. The spread of the drive allele and its effect on population suppression. The figure illustrates a scenario where there is a large selective disadvantage ($s = 0.45$) and transmission effect ($e = 0.95$) that result in strong population suppression (reducing population mean fitness from 1 to 0.55). The figure presents the frequencies of the three alleles (drive allele D in red, nondrive allele d in grey, and resistant allele δ in blue, with corresponding frequencies f_D, f_d and f_δ) and the corresponding value for population mean fitness ($\bar{\omega}$) being indicated by the dashed black line (where $\bar{\omega} = \omega_D f_D + \omega_d f_d + \omega_\delta f_\delta$; with fitness corresponds to the equations in Table 1) across generations. Because population mean fitness is scaled to start at a value of 1 in the absence of the drive allele, the value of $\bar{\omega}$ across generations represents the degree of population suppression. The drive allele is introduced at a 1% frequency ($f_D = 0.01$) in generation (t) 25 and the resistant allele is introduced at a 1% frequency ($f_\delta = 0.01$) in generation 100. The population genetic model used to calculate the allele frequency dynamics follows (Burt, 2003).

Table 1. Fitness of different alleles at a drive locus, arising from different genotype combinations within diploid individuals. Table entries correspond to the fitness of the allele to the left of each row when in a genotype with the allele listed at the top of each column, where fitness can be broadly conceived as the number of gametes of that genotype that are produced. There are three alleles at the drive locus: the drive locus has a drive allele D, which encodes a toxin and antitoxin, and nondrive allele d, and the 'resistant' allele δ, which encodes the antitoxin only. The drive allele encodes a toxin-antitoxin pair which incurs a selective disadvantage $(-s)$ due to toxicity, but can gain a transmission advantage $(+e)$ against a nondrive allele. The resistant allele encodes the antitoxin only, and so is resistant to the drive allele though it still incurs the toxin's selective disadvantage when paired with it. The total fitness of each allele depends on the frequency with which it interacts with each of the three other alleles, which depends on their frequencies in the population.

	D	d	δ
D	$1-s$	$(1+e)(1-s)$	$1-s$
d	$(1-e)(1-s)$	1	1
δ	$1-s$	1	1

We next consider selection on a modifier locus to understand the evolution of resistance to the drive system. For simplicity, we do not consider linkage disequilibrium here (but see SI Appendix, Section 1 for an analysis and justification of ignoring it here). We postulate that modifier activity is constrained such that a modifier allele is only able to influence the expression of a drive allele if they are in the same cell. We consider modifiers of two basic forms: a modifier that achieves suppression by blocking the production of both the toxin and the antitoxin, and a modifier that shows partial suppression by blocking the production of the antitoxin only. Although each of the three forms of killing that we consider is modelled via the same basic one-locus system, they differ from one another with respect to the opportunities for modifiers to alter the outcome of the interaction. This difference arises because the components of hard selection (caused by the toxicity of the drive allele; captured by $-s$) and soft selection (arising from transmission effect; captured by $\pm e$) only apply to the interacting cells among which killing takes place, which varies across the different scenarios.

We first evaluate a drive allele for pre-meiotic killing, where the drive allele is expressed in diploid cells (either parental or somatic). Given that we assume that a modifier can only influence the expression of a drive allele if they are in the same cell, a modifier allele for full or partial suppression has the same effect because the modifier is always in the same cell as all the drive alleles within the individual. Consequently, the modifier is able to block all toxin production, which means that the drive allele is only active in the absence of the modifier. In this way, the modifier has a 'dominant' phenotype, with one allelic copy being able to perform the same role as two allelic copies. With the addition of the modifier (M, with frequency f_M) and nonmodifier (m, with frequency f_m) alleles to form a two-locus system (where there is a drive locus and a modifier locus), the drive (D) and nondrive (d) alleles have fitness (see also SI Appendix, Table S1):

$$\omega_D = 1 + {f_m}^2\left(-s + f_d e(1-s)\right) \tag{3a}$$

$$\omega_d = 1 - f_D {f_m}^2 \big(s + e(1-s)\big) \quad (3b)$$

The condition for the drive allele to have higher fitness than the nondrive allele is the same as before (see eqn. 2), but the two alleles have equal fitness when the modifier allele is at fixation (*i.e.* $f_M = 1$, and hence $f_m = 0$). At the modifier locus, the modifier (M) and nonmodifier (m) alleles have fitness:

$$\omega_M = 1 \quad (4a)$$

$$\omega_m = 1 - (1 - {f_d}^2) f_m s \quad (4b)$$

Consequently, the modifier allele is expected to go to fixation because it always has higher fitness than the nonmodifier allele, though the two alleles trivially have equal fitness when the drive allele is extinct (*i.e.* $f_D = 0$). Therefore, a drive allele for pre-meiotic killing is susceptible to a modifier for full (and/or partial) suppression.

We can next examine a drive allele for post-meiotic killing, where the drive allele is expressed in the haploid daughter cells (*i.e.* gametes) after meiosis has taken place. We first consider a modifier that fully suppresses the drive allele by blocking both toxin and antitoxin production. In this scenario, the modifier has no control over the expression of the toxin and antitoxin within the gametes that do not carry the modifier. Consequently, for an individual that is heterozygous at the drive and modifier loci (*i.e.* the diploid genotype is $DdMm$), the modifier allele blocks the expression of the toxin and antitoxin in 25% of the gametes but another 25% of the gametes still produce the toxin and antitoxin without being blocked by the modifier. Therefore, the modifier has a 'recessive' phenotype, as an individual has to be homozygous for the modifier to block toxin production in all gametes. In this scenario (see SI

Appendix, Table S2A), the fitness of each allele in the two-locus two-allele system is given by:

$$\omega_D = 1 + \left(1 - {f_M}^2\right)\left(-s + f_d e(1-s)\right) \quad (5a)$$

$$\omega_d = 1 - f_D\left(1 - {f_M}^2\right)\left(s + e(1-s)\right) \quad (5b)$$

$$\omega_M = 1 - f_m\left(1 - {f_d}^2\right)\left(s + e(1-s)\right) \quad (5c)$$

$$\omega_m = 1 + \left(1 - {f_d}^2\right)\left(-s + f_M e(1-s)\right) \quad (5d)$$

Note the symmetries between the fitness equations of alleles D (drive allele) and m (nonmodifier allele), and between alleles d (nondrive allele) and M (modifier allele), which captures the nature of selection in this system. The modifier allele receives the same fitness as the nondrive allele because the modifier allele exhibits the phenotype of the nondrive allele; the modifier allele prevents the expression of the drive allele, and so avoids producing its toxicity costs whilst also making itself susceptible to drive. Concurrently, the fitness of the drive and nonmodifier allele are interdependent and are each favoured when:

$$f_D f_m > \frac{1}{s}\left(1 - (1+e)(1-s)\right) \quad (6)$$

This condition recapitulates the condition where the drive allele is favoured in a one-locus system (eqn. 2) when the nonmodifier allele is at fixation ($f_m = 1$) but, as the frequency of the nonmodifier allele declines, the frequency above which the drive allele is favoured increases. Due to this interdependence, the conditions that favour the drive allele also favour

the nonmodifier allele (and vice versa). Thus, we expect the drive allele to go to fixation and the modifier allele to go extinct.

Staying with the scenario of post-meiotic killing, we can next consider the case where a modifier allele partially suppresses the drive allele by only blocking toxin production (whilst not blocking antitoxin production). There are only two diploid genotypes ($DDMm$ and $DdMm$) in which the drive allele has different fitness under partial and full suppression by modifiers, where a fully suppressing modifier allele is the victim of transmission distortion whilst a partially suppressing modifier allele avoids this fate. Within these different interactions (see SI Appendix, Table S2B), there is no change to the fitness of alleles at the drive locus (see eqn. 5) and so the conditions favouring the drive allele are the same as for full suppression (eqn. 6), but the fitness of alleles at the modifier locus are changed:

$$\omega_M = 1 - (1 - {f_d}^2) f_m s \tag{7a}$$

$$\omega_m = 1 - (1 - {f_d}^2) s \tag{7b}$$

Consequently, the modifier allele is expected to go to fixation because it always has higher fitness than the nonmodifier allele (except the trivial case where the drive allele is extinct, *i.e.* $f_D = 0$). Therefore, a drive allele for post-meiotic killing is not susceptible to a modifier allele that causes full suppression, but it is susceptible to a modifier allele that causes partial suppression.

Lastly, we examine the scenario of meso-meiotic killing, where the drive allele is expressed in sister cells after the first reductional meiotic division and before the second equational meiotic division. We first consider a modifier allele that fully suppresses the drive allele by blocking both toxin and antitoxin production. As with the post-meiotic killing scenario, these modifier alleles that cause full suppression of the drive allele have no control

over the expression of the toxin and antitoxin within the gametes that do not carry the modifier. The meso-meiotic and post-meiotic killing scenarios only differ in the pattern of allelic fitness in the double-heterozygote individuals ($DdMm$; see SI Appendix, Figure S3) because, in all other scenarios, the two sister cell genotypes are the same as the four daughter cell genotypes; in the double-heterozygotes, post-meiotic killing takes place between all four daughter cells (which have all possible haploid genotypes) but meso-meiotic killing takes place between sister cells that always have the opposite genotype. Thus, in the double-heterozygotes, toxin production only takes place in one sister cell pair (DmxdM; not DMxdm), limiting which gametes suffer the selective disadvantage ($-s$) and transmission effect ($\pm e$). In this way (see SI Appendix, Table 3SA), the fitness of each allele in the two-locus two-allele system is given by:

$$\omega_D = 1 + f_m\big(-s + f_d e(1-s)\big) - f_D f_M f_m s \tag{8a}$$

$$\omega_d = 1 - f_D f_m\big(s + e(1-s)\big) \tag{8b}$$

$$\omega_M = 1 - f_D f_m\big(s + e(1-s)\big) \tag{8c}$$

$$\omega_m = 1 + f_D\big(-s + f_M e(1-s)\big) - f_D f_d f_m s \tag{8d}$$

Again, the drive allele and the nonmodifier are favoured under the same conditions as for the case of post-meiotic killing (see eqn. 6). Further, similarly to modifier alleles that cause full suppression under post-meiotic killing, there are symmetries between the fitness of the drive allele (D) and nonmodifier allele (m), and between the nondrive allele (d) and the modifier allele (M) which capture the nature of selection in the system.

We can next consider the case where a modifier allele partially suppresses the drive allele by only blocking toxin production. In the scenario with meso-meiotic killing, partial suppression is very similar to full suppression except, when the drive allele is homozygous and the modifier allele is heterozygous ($DDMm$), there is no drive despite there being toxin production (by the Dm sister cell) because both sister genotypes produce the antitoxin. Consequently, the fitness of the alleles at the drive locus are unchanged from the full suppression scenario (see eqn. 8; see also SI Appendix, Table S3B), but the fitness of the alleles at the modifier locus are:

$$\omega_M = 1 - f_D f_m\big(s + f_d e(1-s)\big) \tag{9a}$$

$$\omega_m = 1 + p\big(-s + f_d f_M e(1-s)\big) - f_D f_d f_m s \tag{9b}$$

Given that the fitness of alleles at the drive locus are the same as in the case where modifiers cause full suppression, the drive allele is favoured under the same condition as it is in the scenario with post-meiotic killing (see eqn. 6). But the slightly different expressions for fitness of alleles at the modifier locus mean that the modifier allele is favoured under a slightly different condition:

$$f_M > \frac{-f_d}{f_D s}\big(1 - (1+e)(1-s)\big) \tag{10}$$

Given that the condition $(1+e)(1-s) > 1$ has to be met for the drive allele to invade, the RHS is positive for all relevant parameter space (see SI Appendix, Figure S2). Consequently, a modifier allele that blocks toxin production must exceed a threshold frequency to invade the nonmodifier allele (though once it has exceeded this threshold, it is expected to spread to

fixation). The threshold frequency declines as the drive allele's frequency increases to the point where the modifier is always able to invade when the drive allele is at fixation ($f_D = 1$). The reason for this frequency threshold is the fitness asymmetry when both the drive and the modifier loci are heterozygous (*i.e.* the diploid genotype is $DdMm$): a modifier allele (M) can lose out from unfair meiosis when it is in a gamete with a nonmodifier allele (m), whilst it blocks toxin production to restore fair meiosis when the situation is reversed. As a result, a gene drive by meso-meiotic killing can resist the spread of partially suppressing modifiers that block toxin production across most relevant parameter values (*i.e.* when $(1 + e)(1 - s) > 1$ and $0 \leq f_D < 1$). Moreover, this means that the conditions that promote invasion by the drive allele (small s and large e) also prevent the invasion of the modifier that blocks toxin production.

Although a gene drive via meso-meiotic killing can evade the evolution of resistance in the shorter-term (arising from either fully or partially suppressing modifiers), the evolution of resistance in the longer term seems likely because a resistant allele could arise at the drive locus through a mutation that renders the toxin non-functional (whilst retaining a functional antitoxin). As applies for all three types of killing we have discussed, the resistant allele would always be favoured because the drive allele would never get the chance to reap a transmission advantage (e) against it. In contrast to a partially suppressing modifier under meso-meiotic killing, a resistant allele is favoured whilst a resistant-creating modifier allele is not because the resistant allele does not suffer the asymmetric drive outcomes in the diploid genotype $DdMm$ (see above). Further, because the resistant allele would not produce a functional toxin, it would not suffer the selective disadvantage ($-s$) when it is homozygous. Thus, when the resistant allele (δ) competes against the other alleles at that locus (the drive D and nondrive d alleles), their fitness can be expressed as (Table 1):

$$\omega_D = 1 - s + f_d e(1 - s) \quad (11a)$$

$$\omega_d = 1 - f_D s - f_D e(1 - s) \quad (11b)$$

$$\omega_\delta = 1 - f_D s \quad (11c)$$

Thus, the resistant allele (δ) always has higher fitness than the nondrive allele (d), and is expected to go to fixation because the resistant allele (δ) also has higher fitness than the drive allele (D) whenever the nondrive allele (d) has a low frequency. However, the resistant allele (δ) only has a marginally higher fitness when the drive allele (D) is near fixation, and so is expected to spread to fixation slowly (compared to the drive allele's rapid spread to fixation against the nondrive allele; see Figure 2). In this way, there is an unavoidable vulnerability of a gamete-killing gene drive to resistant alleles, which seem likely to arise in the longer-term.

In summary, we have analysed the impact that the timing of when a gamete-killing drive allele is active during meiosis has on selection for modifiers at other loci that can fully or partially suppress the drive allele (and thereby generating resistance to the drive allele). We propose that meso-meiotic killing provides the best timing for a drive allele to be active because a meso-meiotic killing drive allele is invulnerable to both full and partial suppression by modifiers. In this way, the meso-meiotic killing drive allele can evade the rapid evolution of resistance in the shorter-term. However, in the longer-term, all forms of gamete-killing (including meso-meiotic killing) are susceptible to a resistant allele at the drive allele's locus. Whilst this might seem to limit the utility of transmission distorting gene drives by meso-meiotic killing, susceptibility to a resistant allele can be potentially desirable because it makes the gene drive system controllable. For example, a resistant allele could be introduced to limit the spread of a drive allele beyond a target population or to reverse a population suppression were it to have unintended consequences. A limitation of this control is that the population becomes immune to the drive allele (in as far as the resistant allele persists), and so a new

drive construct would need to be created to attempt any further population controls (*e.g.* a new toxin-antitoxin pair).

Beyond the problem of improving gene drive efficacies, meso-meiotic killing may explain the surprisingly long-term persistence (or stability) of well-documented meiotic drive systems despite the potential for strong selection of resistance (Lindholm *et al.*, 2016). For example amongst the well-studied autosomal drives, *segregation distorter* (*SD*) in *Drosophila melanogaster* has persisted for over 1 million years (Kovacevic & Schaeffer, 2000), the *t*-haplotype in *Mus musculus* has persisted over 2 million years (Silver, 1993) and *spore killer* (*Sk*) in *Neurospora intermedia* may have persisted for up to 0.5 million years (Svedberg *et al.*, 2018). As it stands, the mechanistic details about how these meiotic drives function remains incomplete, but meso-meiotic killing presents an interesting hypothesis that when these drive alleles become active during meiosis could be an important part of explaining their surprising evolutionary longevity. The critical feature of this hypothesis is that the drive occurs during the interaction between the pair of cells that have just undergone reductive division, and so for species that have 'inverted meiosis' (where the reductive division takes place after an equational division) the same hypothesis can still be considered.

When looking for plausible systems with the profile of meso-meiotic killing, the many examples of meiotic drive that are known to operate through the disruption of the equational division could be plausible cases, like the Paris X^{SR} chromosome which disrupts heterochromatin during anaphase II (Cazemajor *et al.*, 2000) – though this system has numerous differences to the one modelled here. The Paris X^{SR} chromosome only operates in males during the formation of sperm. Although we have not modelled gamete-killing taking place within a specific sex, there is in principle no change to the selection on modifiers in as far as the sex ratio is not distorted. Sperm-killing may be more common than means of gamete-killing than egg-killing because, in many taxa (including *Drosophila* for the Paris X^{SR} chromosome system), meiotic products are compartmentalised during spermatogenesis whilst they are not during oogenesis, giving greater potential for targeted killing. However, the Paris

X^{SR} chromosome does distort the sex ratio, which gives greater potential for selection to favour both fully and partially suppressing modifiers under broader conditions to restore an equal sex ratio. Additionally, meso-meiotic killing requires expression in the sister cells, which is not thought to be the case for the Paris sex ratio distorter where there is evidence that the killing protein is expressed and binds the Y-chromosome (leading to its eventual destruction in anaphase II) pre-meiotically (Helleu *et al.*, 2016); the Paris X^{SR} chromosome is one of the few systems where the critical expression data has been collected. Although the meso-meiotic killing hypothesis fails to describe this case, the general mechanism of the Paris X^{SR} chromosome's meiotic drive in heterochromatin disruption is a possible pathway to targeted killing between sister cells. Furthermore, speculating in the general vacuum of relevant evidence either way, as meso-meiotic killing has equivalent results to post-meiotic killing, many examples that are assumed to reflect post-meiotic killing may actually be of this meso-meiotic form.

Acknowledgements

We would like to thank our funders: SWBio DTP (BBSRC) for P. G. M. and BBSRC for J. B. W. (BB/M01035X/1). We would also like to thank Austin Burt and Mike Wade for comments on an earlier version of the manuscript.

References

Beeman, R.W., Friesen, K.S. & Denell, R.E. 1992. Maternal-effect selfish genes in flour beetles. *Science (80-.).* **256**: 89–92.

Bravo Núñez, M.A., Nuckolls, N.L. & Zanders, S.E. 2018. Genetic Villains: Killer Meiotic Drivers. *Trends Genet.* **34**: 424–433.

Buchman, A., Marshall, J.M., Ostrovski, D., Yang, T. & Akbari, O.S. 2018. Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proc. Natl. Acad. Sci.* **115**: 4725–4730.

Bull, J.J. & Malik, H.S. 2017. The gene drive bubble: New realities. *PLoS Genet*. **13**: e1006850.

Burt, A. 2014. Heritable strategies for controlling insect vectors of disease. *Philos. Trans. R. Soc. B Biol. Sci.* **369**: 20130432.

Burt, A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. B Biol. Sci.* **270**: 921–928.

Burt, A. & Trivers, R. 2008. *Genes in Conflict*. Harvard University Press.

Cazemajor, M., Joly, D. & Montchamp-Moreau, C. 2000. Sex-ratio meiotic drive in Drosophila simulans is related to equational nondisjunction of the Y chromosome. *Genetics* **154**: 229–236.

Champer, J., Buchman, A. & Akbari, O.S. 2016. Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**: 146–159.

Champer, J., Liu, J., Oh, S.Y., Reeves, R., Luthra, A., Oakes, N., et al. 2018. Reducing resistance allele formation in CRISPR gene drive. *Proc. Natl. Acad. Sci.* **115**: 5522–5527.

Champer, J., Reeves, R., Oh, S.Y., Liu, C., Liu, J., Clark, A.G., et al. 2017. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **13**: 1–18.

Charlesworth, B. & Hartl, D.I. 1978. Population dynamics of the segregation distorter polymorphism of Drosophila melanogaster. *Genetics* **89**: 171–192.

Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L. V., Guo, M., et al. 2007. A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science (80-.).* **316**: 597–600.

Deredec, A., Burt, A. & Godfray, H.C.J. 2008. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**: 2013–2026.

Godfray, H.C.J., North, A. & Burt, A. 2017. How driving endonuclease genes can be used to combat pests and disease vectors. *BMC Biol.* **15**: 1–12.

Haig, D. & Grafen, A. 1991. Genetic scrambling as a defence against meiotic drive. *J. Theor. Biol.* **153**: 531–558.

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., et al. 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**: e1007039.

Hartl, D.L. 1975. Modifier theory and meiotic drive. *Theor. Popul. Biol.* **7**: 168–174.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B., et al. 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proc. Natl. Acad. Sci.* **113**: 4110–4115.

Hornett, E.A., Charlat, S., Wedell, N., Jiggins, C.D. & Hurst, G.D.D. 2009. Rapidly Shifting Sex Ratio across a Species Range. *Curr. Biol.* **19**: 1628–1631.

Kovacevic, M. & Schaeffer, S.W. 2000. Molecular population genetics of X-linked genes in Drosophila pseudoobscura. *Genetics* **156**: 155–172.

Leigh, E.G. 1971. *Adaptation and Diversity: natural history and the mathematics of evolution*. Freeman, Cooper.

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., et al. 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol.* **31**: 315–326.

Lyttle, T.W. 1977. Experimental population genetics of meiotic drive systems. I. Pseudo Y chromosomal drive as a means of eliminating cage populations of Drosophila melanogaster. *Genetics* **86**: 413-445.

Marshall, J.M. & Akbari, O.S. 2018. Can CRISPR-Based Gene Drive Be Confined in the Wild? A Question for Molecular and Population Biology. *ACS Chem. Biol.* **13**: 424–430.

McFarlane, G.R., Whitelaw, C.B.A. & Lillico, S.G. 2018. CRISPR-Based Gene Drives for Pest Control. *Trends Biotechnol.* **36**: 130–133.

McLaughlin, R.N. & Malik, H.S. 2017. Genetic conflicts: the usual suspects and beyond. *J. Exp. Biol.* **220**: 6–17.

Noble, C., Adlam, B., Church, G.M., Esvelt, K.M. & Nowak, M.A. 2018. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *Elife* **7**: e33423.

Noble, C., Olejarz, J., Esvelt, K.M., Church, G.M. & Nowak, M.A. 2017. Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3**: 3–10.

Palopoli, M.F. & Wu, C. 1996. Rapid Evolution of a Coadapted Gene Complex: Evidence From the Segregatuib Distorter (SD) System of Meiotic Drive in Drosophila melanogaster. *Genet. Soc. Am.* **143**: 1675–1688.

Pinzone, C.A. & Dyer, K.A. 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of Drosophila neotestacea. *Proc. R. Soc. B Biol. Sci.* **280**: 20131397.

Price, T.A.R., Hurst, G.D.D. & Wedell, N. 2010. Polyandry Prevents Extinction. *Curr. Biol.* **20**: 471–475.

Prout, T., Bundgaard, J. & Bryant, S. 1973. Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theor. Popul. Biol.* **4**: 446–465.

Sandler, S. & Novitski, E. 1957. Meiotic drive as an evolutionary force. *Am. Nat.* **857**: 105–110.

Silver, L.M. 1993. The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet.* **9**: 250–254.

Svedberg, J., Hosseini, S., Chen, J., Vogan, A.A., Mozgova, I., Hennig, L., et al. 2018. Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nat. Commun.* **9**: 4242.

Thomson, G.J. & Feldman, M.W. 1976. Population genetics of modifiers of meiotic drive III. Equilibrium analysis of a general model for the genetic control of segregation distortion. *Theor. Popul. Biol.* **10**: 10–25.

Unckless, R.L., Clark, A.G. & Messer, P.W. 2017. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**: 827–841.

Wade, M.J. & Beeman, R.W. 1994. The population dynamics of maternal-effect selfish genes. *Genetics* **138**: 1309–1314.

Comment 5: From greenbeards to unbeatable genes

Inspired by newly-discovered greenbeard genes, the preceding paper provides an exploration of how to design gene drives in a way that avoids resistance from modifier genes. Here, I develop the core logic to present the concept of an unbeatable gene.

Under natural selection, we expect genes to become proficient at maximizing their replication into the next generation (Dawkins, 1976). The most proficient genes are undoubtedly selfish genetic elements, which pillage the organism to maximize their replication, often at the expense of the replication of other genes in the rest of the genome (Werren *et al.*, 1988). However, as replication necessarily invokes sociality, a successful gene would do well to work together with copies of itself against its competitors (Hamilton, 1972). Therefore, at its best, a gene in an actor would conditionally cooperate with a recipient only if it shared a copy of that gene (Hamilton, 1964b, 1972). Such a gene is a greenbeard, changing its behaviour dependent on whether or not others share a copy of the greenbeard gene (Hamilton, 1964b; Dawkins, 1976). Yet, as Paper 4 clarified (see Table 1 in Paper 4), it has been the general expectation that greenbeard genes are vulnerable to modifiers for full/partial suppression or falsebeards that cheat the greenbeard signal. In this way, not all greenbeards are going to be 'unbeatable genes' because, although they are good competitors against other alleles, many greenbeards are likely to be poor competitors against genes at other loci. The claims of greenbeard discoveries might suggest that some greenbeards have found ways around the problem of modifiers, but it remains unclear how any real greenbeards might do this. It seems likely that the biological constraints of the particular systems are important, just as the structure of the interaction between cells during meiosis is important in Paper 5.

As a definition: an unbeatable gene is a gene that can outperform all its allelic competitors whilst also being immune to modification by genes at other loci. In practical terms, I would make the case that an unbeatable gene would be a selfish genetic element because these genes have an elevated rate of transmission compared to other genes at a locus.

Further, I would also claim that there is an intimate (if not exclusive) relationship between an unbeatable gene and a greenbeard because the social interaction between different individuals that carry genes provides a way in which the gene can escape modification. Although not explicitly addressed, the idea that sociality provides an escape from modification is found within Paper 5, which showed that a meiotic drive gene can distort its transmission into the next generation whilst also escaping modification. All the killing phenotypes can be conceptualised as forms of social interaction between cells with different genotypes, which is the link to greenbeards. There are parallels between Hamilton's (1964b) cost-free concept of a greenbeard and the pre-meiotic killer because both are susceptible to fully suppressing modifiers that are kin-selected based on a different background level of relatedness between the drive and modifier allele. Further, both concepts give no distinction between a modifier for full or partial suppression. There are also parallels between Dawkins' (1976) indirect fitness concept of a greenbeard and the post-meiotic killer because both are exclusively vulnerable to a partially suppressing modifier which creates a (falsebeard) cheater. Further, there is no intragenomic conflict in either case because both are also vulnerable to cheater alleles. Finally, Haig's (1996) meiotic drive concept of a greenbeard and the meso-meiotic killer share many similarities. Although Haig (1996) directly states that a meiotic drive-like greenbeard would be vulnerable to modifiers, the direct analogy between meiotic drive and greenbeards does implicitly identify the critical idea that social interaction has a special property for immunity to modification because modifiers' effects cannot necessarily access all the interacting genotypes. Consequently, the meso-meiotic killer has some resistance to modifiers because of the structure of the social interaction among sister cells (see Figure 1 in Paper 5).

The meso-meiotic killer is not an unbeatable gene for two reasons, which may reduce its utility as a means of gene drive. First, the drive allele for meso-meiotic killing is vulnerable to partially suppressing modifiers when it is at fixation. In principle, from the perspective of gene drive design, there may be ways of designing the toxin-antitoxin system to negate (or, at

least, make harder to evolve) the possibility of a partially suppressing modifier, such as having a single promoter for the toxin and antitoxin genes. Second, the drive allele for meso-meiotic killing is vulnerable to a resistant (*i.e.* falsebeard) allele that only produces the antitoxin, and so is not able to outcompete all of its allelic competitors. This point was advanced as a beneficial feature because it means that the gene drive could be eliminated if it had unintentional consequences. For this reason, gene drive technologies would usually do well to evade resistance, but not permanently escape it, as a hypothetically unbeatable gene would do. Nonetheless, there are purposes (*e.g.* population eradication) where an unbeatable gene may be preferred over a more controllable gene that is vulnerable to resistant alleles.

Although the question of how unbeatable genes could work remains wide open, they are interesting to speculate on, especially in the context of gene drive where they have a practical purpose. Herein, the language of modifiers provides a way of talking about how genes at multiple loci coevolve together (Leigh, 1971; Hurst *et al.*, 1996). In its historical usage, a modifier for almost any effect has been discussed (*e.g.* effecting mutation rate, recombination rate, suppressing or enhancing a phenotype *etc*), but to avoid being impractically speculative is important to discuss modifiers in a disciplined way. In Paper 5, using a genetic construct of a toxin and antitoxin, of which there are numerous possible biochemical mechanisms (Burt & Trivers, 2008; Bravo Núñez *et al.*, 2018), I considered fully suppressing modifiers that either prevent toxin and antitoxin production or partially suppressing modifiers that only prevent toxin production whilst permitting antitoxin production – which are both mechanistically reasonable. Consequently, I would suggest that meso-meiotic killing gene drive comes the closest to presenting a realistic gene that could be unbeatable, even though it falls short of the mark.

Discussion

My thesis makes the case that conditional cooperation is likely to be a more important part of the evolutionary explanation of cooperation in nature than current evidence might suggest, using a mixture of theoretical and experimental arguments. In the Introduction, I exposed how there has been a great deal of confusion surrounding this claim, which has been confounded with issues relating to the conflation of altruism and cooperation, the over-emphasis of the problem of cheating, and both theoretical and experimental attention focusing on facultative forms of conditionality. Setting these issues aside, there are two primary sources of scepticism about the role that conditional cooperation plays in the broader evolutionary explanation of cooperative behaviour: first there is little evidence for conditional cooperation, and second it is unclear through what mechanism conditional cooperation could operate. Whilst my thesis presents some evidence to rebut the first claim, it is beyond the scope of this thesis to comprehensively demonstrate any general rebuttal given the general absence of evidence. Nonetheless, starting from the observation that different strains of *Dictyostelium discoideum* appear to exhibit condition cooperation when mixed at different frequencies (Buttery *et al.*, 2009), Paper 1 generated game theoretic predictions for how chimeric mixes of two strains are expected to vary their level of cooperation across frequencies, and then tested these predictions using naturally co-occurring strains to show a strikingly close fit of theory and data. Paper 1 also sought to explain the predicted patterns of conditional cooperation with appeal to the familiar logic of the prisoner's dilemma and the snowdrift games, which were explored in more detail in Comment 1. Across frequencies, strains with near-equal frequencies experienced a social dilemma that was more similar to the prisoner's dilemma, whilst strains with more unequal frequencies experienced a social dilemma that was more similar to the snowdrift. In Comment 1, I went into more detail to calculate the boundary condition (in the 'snowdrift rule'), which clarified how a strain's level of cooperation is self-interested – to ensure that the strain reaps the majority (>50%) of the benefit from its investment.

Although the fit between predictions and data was striking (as a rebuttal to the first source of scepticism), Paper 1 provided no details of the mechanism through which strains were adjusting their level of cooperation in different social situations (*i.e.* without providing any answers to the second source of scepticism). Paper 2 examined the predictions in chimeric mixes with more than two strains, confirming the general patterns of conditional cooperation for chimeras of two strains (from Paper 1) for chimeras of three strains. However the predictions, which are made assuming that strains have perfect information about the investment of other strains, do not accurately predict how strains should be uncooperative for groups with a large number of players (even after adjustment for imperfection information about each strains' relatedness to the group). As Comment 2 examines further, strains do not appear to be able to detect the number of competing strains they are in a chimera with, which is suggestive of the mechanism through which they detect their relatedness to the group. This inability to detect the number of competing strains suggests that a self-binding signal-receptor may be providing the information (like TgrB1/C1), creating a nonadaptive constraint that limits how uncooperative strains will be. This opens up the question of how any 'self' signalling system could maintain the polymorphism needed for successful genetic recognition (known as Crozier's paradox). Paper 3 explores a potential explanation for Crozier's paradox by examining the underlying game theoretic model in greater detail to explain the logic of conditional cooperation. Surprisingly, for any number of players (like strains), the model shows that higher relatedness players have lower fitness because of conditional cooperation: higher relatedness players contribute more toward cooperation giving a disproportionate advantage to lower relatedness players – akin to negative frequency-dependent fitness, which provides a resolution to Crozier's paradox. The model is abstract, and so Comment 3 examines how robust this result is to real-world mechanisms. The property of negative frequency dependence is shown to be a fundamental prediction of the form of conditional cooperation we consider. In this way, Papers 2 and 3 provide some clarity as to the plausible mechanisms that can sustain conditional cooperation under a broad range of conditions.

To examine one potentially plausible mechanism of conditional cooperation, Paper 4 examines the recent surge of experimental evidence for greenbeard genes across study systems. Greenbeard genes, which rely on conditional social behaviour, have been dismissed as unimportant to the evolutionary explanation of social behaviour. However, numerous examples have received considerable empirical support in the last few years, though whether or not they are really greenbeards remains contentious (as they may be kin recognition genes). Either way (whether greenbeard or kin recognition), these study systems provide a molecular mechanism for conditional cooperation, showing the evolutionary potential to underlie conditional cooperation through their (often extreme) polymorphism. Of these examples, it is unclear whether or not adjustment of the level of cooperation could help to explain their polymorphism, but Paper 1 presented evidence that the segregation behaviour in *D. discoideum* that is controlled by *tgrb1*/*tgrc1* greenbeard (Gruenheit *et al.*, 2017) varies both adaptively and quantitatively across different frequencies in chimera. As Paper 4 pointed out, *tgrb1*/*tgrc1* is probably the most plausible of the greenbeard (not kin-selected) genes because there has been the critical demonstration that segregation correlates to relatedness at the *tgrb1*/*tgrc1* locus rather than genome-wide relatedness (Gruenheit *et al.*, 2017). This claim is bolstered because *tgrb1*/*tgrc1* locus is one of very few genes (among other members of the *tgr* gene family) with high enough polymorphism to explain the partner specificity of segregation behaviour that has been observed among the test-set of naturally co-occurring strains (Gruenheit *et al.*, 2017; de Oliveira *et al.*, 2019). For a similar reason, it seems likely that *tgr*-like genes are also responsible for the partner specificity of relatedness-dependent investment behaviour – though the details about how this could work remain obscure. Therefore, although the observed behaviour from Paper 1 and 2 is suggestive of a polymorphic signal-receptor, and Paper 3 suggests that such polymorphism could be maintained by conditional cooperation in this system, and Paper 4 hints that a greenbeard gene may be responsible, the quest for the gene that causes conditional cooperation through stalk investment behaviour in *D. discoideum* remains to be discovered.

Moving on from *D. discoideum*, Comment 4 clarifies the logic of different presentation of the greenbeard thought experiment using some simple models, showing how different conceptions of greenbeard genes make different predictions about their properties. Greenbeard genes have variously been described to draw attention to their role in changing the relatedness between social partners, causing a social behaviour and being selfish genetic elements that can harm individual fitness. All of these conceptions of greenbeards are possibilities that highlight different aspects of the fundamental greenbeard concept (see Paper 4). Even with this diversity of views, there are many outstanding questions about greenbeard genes that provide exciting areas for future research. One of the most intriguing aspects of newly-discovered greenbeard genes is their role in sophisticated 'fusion' behaviours that involve the concerted action of many other genes. For these social behaviours, the greenbeard takes on the role of the signal gene that determines the recipient of the behaviour and has little causal role in the production of the behaviour itself. Given that the behaviour is a polygenic trait, it would seem eminently plausible for the genes involved in the production of the social behaviour to act as modifiers to suppress the greenbeard, but observationally this is not the case. How greenbeard genes might escape modification is fascinating, and likely to be related to the particular biological constraints at play. To consider this idea in greater detail in a more applied concept, Paper 5 took inspiration from greenbeard genes to consider how gene drives could be designed to reduce the risk of modifiers suppressing them (and preventing their population suppressing effects). Using the logic of conditional cooperation to consider gamete killers, Paper 5 suggests how the risk of resistance can be reduced if the killing takes place among sister cells during meiosis (*i.e.* meso-meiotically) because of the particular structure of the cellular interaction. Starting from the greenbeard concept, Comment 5 develops the concept of an unbeatable gene as a design objective for gene drives, suggesting that meso-meiotically killing is amongst the closest to this ideal.

Therefore, starting from a puzzling observation, this thesis presents the case that conditional cooperation is indeed an important part of the evolutionary explanation of

cooperation. I have presented evidence for conditional cooperation through cellular differentiation in *Dictyostelium discoideum*, and suggested a plausible mechanism for conditional cooperation through the *tgr* genes. Further, I have shown that conditional modulation of the level of cooperation can be sufficient to maintain the polymorphism needed for conditional cooperation to persist (*contra* Crozier's paradox). Lastly, I have examined conditional cooperation in a broader context to consider other plausible systems with the hallmarks of conditional cooperation and how understanding these systems could provide useful applications through the improvement of gene drives. In this way, I would claim that conditional cooperation is likely to be an important part of the evolutionary explanation of cooperation and should not be dismissed on the grounds of limited evidence or a lack of plausible mechanisms – because I have shown that evidence can be found when it is carefully looked for and that there are numerous study systems with the hallmarks of conditional cooperation. Furthermore, by considering a useful application of the evolutionary logic of conditional cooperation, I have shown that conditional cooperation is all the more important for its practical utility.

Supplementary Information (SI) Appendix

Paper 1

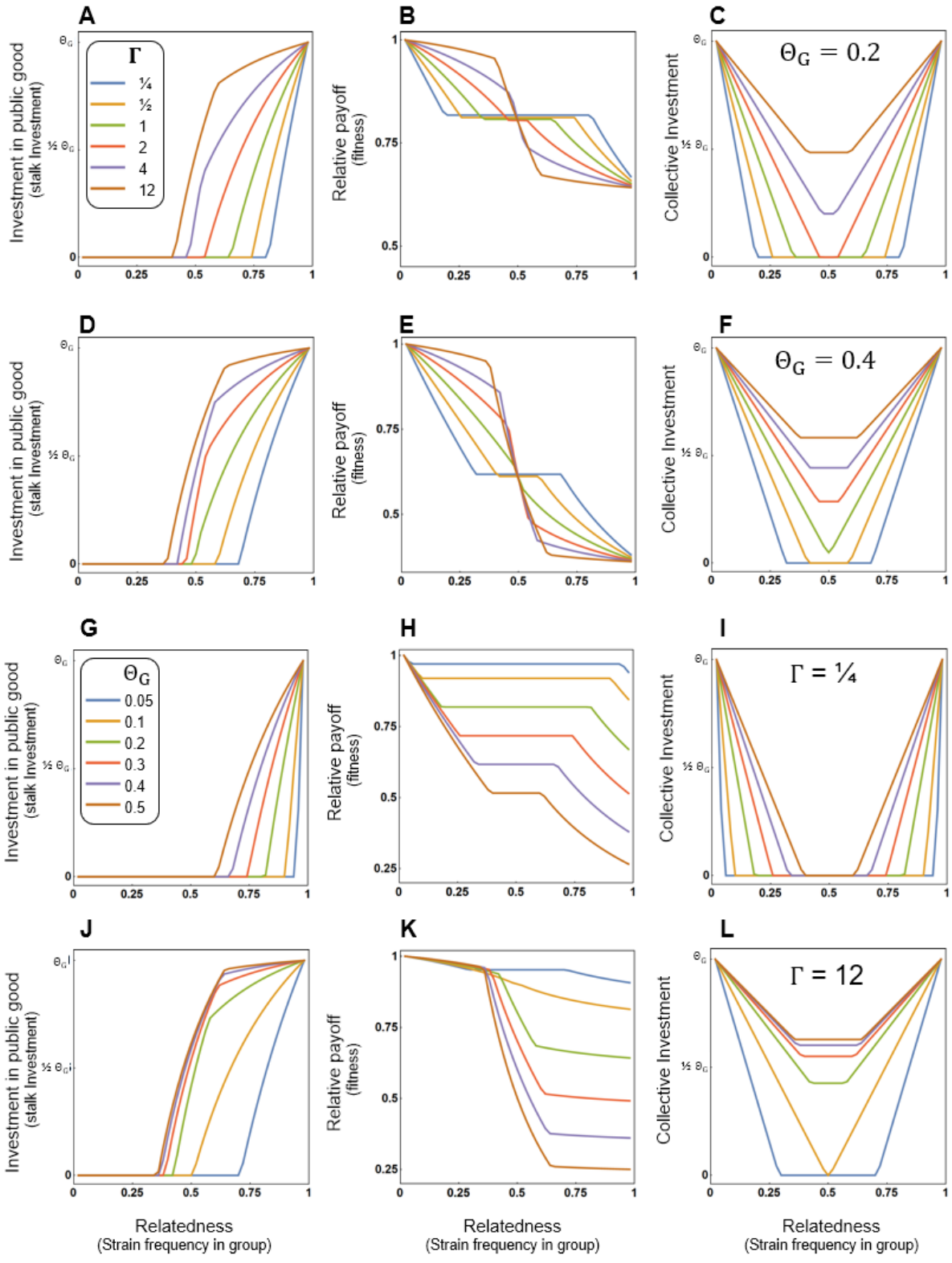


Figure S1. Patterns of individual investment, relative payoffs and collective investment for different optimal levels of collective investment (Θ_G) and strengths of selection (Γ). Each of the first two rows (parts **A** to **F**) show the same relationships as in Figure 2, but for two different optimal levels of collective investment (Θ_G = 0.2 for **A** to **C** and Θ_G = 0.4 for **D** to **F**), with lines in each figure showing the pattern expected for varying strengths of selection (Γ) (which essentially cover the entire possible range of parameter space from very weak to very strong selection on investment; the legend imbedded in the first figure of each row gives the line color of each strength of selection, with all nine panels using the same color coding). The last two rows (parts **G** to **L**) show these same relationships for two different strengths of selection (Γ= ¼ for **G** to **I** and Γ = 12 for **J** to **L**), with the lines in each figure showing the pattern for different optimal levels of collective investment. The labels in parentheses relate the figures to the patterns expected for the *D. discoideum* system. The first figure in each row (**A**, **D**, **G** and **J**) shows the pattern of individual investment into the public good (stalk investment) as a function of the player's relatedness to (frequency in) the group. The second figure in each row (**B**, **E**, **H**, and **K**) shows the pattern of relative payoff (fitness) as a function of the player's relatedness to (frequency in) the group. The last figure in each row (**C**, **F**, **I**, and **L**) shows the pattern of collective investment as a function of a focal player's relatedness to (frequency in) the group.

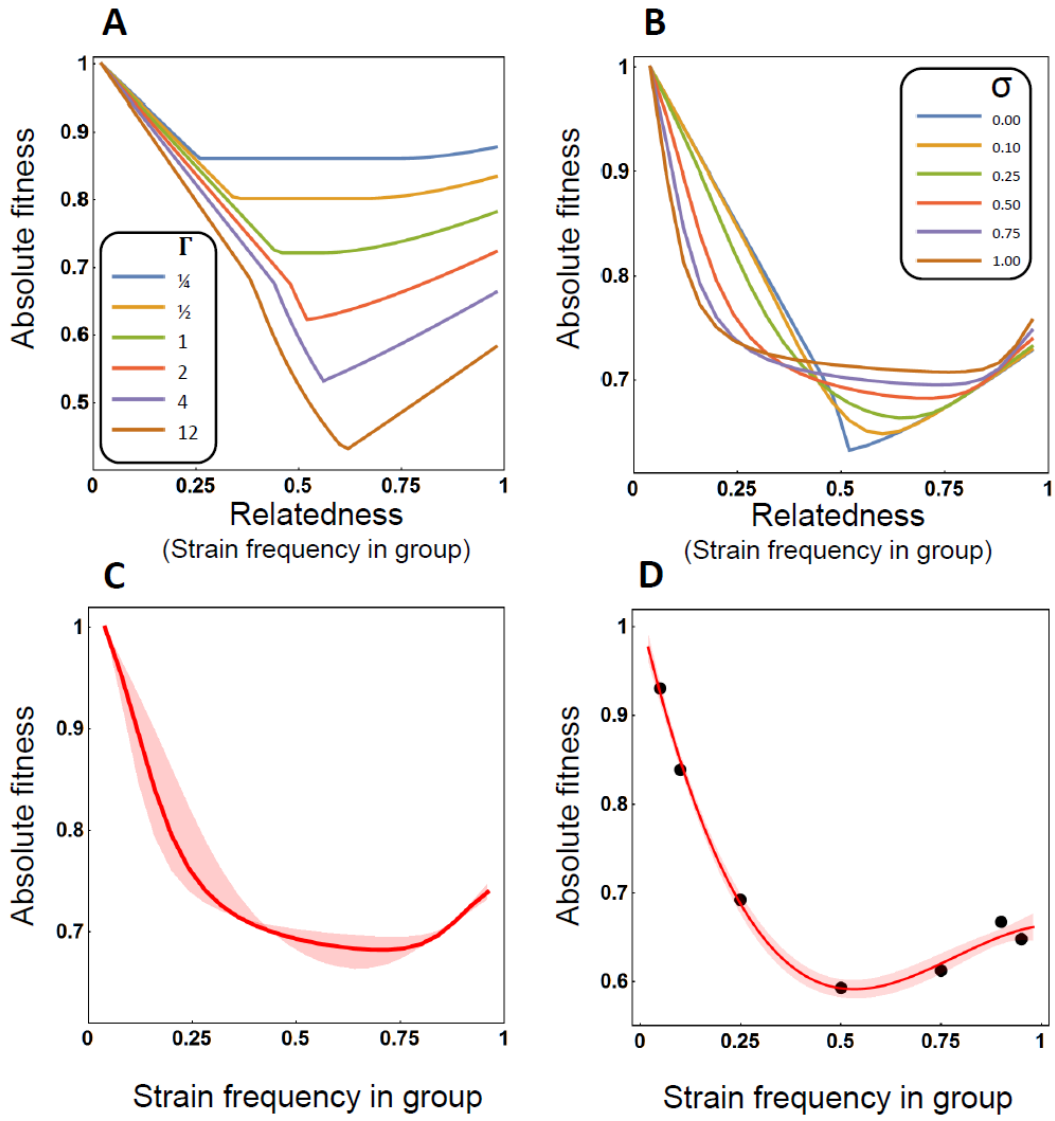


Figure S2. Patterns of absolute fitness as a function of relatedness (frequency in a group). **A)** Patterns of fitness (which represents the total payoff) as a function of relatedness (frequency in a group) when individuals show the ESS pattern of investment in the public good. All patterns were calculated for the same optimal level of investment ($\Theta_G = 0.3$), with the different lines corresponding to different strengths of selection (Γ). **B)** Patterns of absolute fitness as a function relatedness (frequency) with varying degrees of error in measurement of relatedness (frequency). All patterns were calculated for the same optimal level of investment ($\Theta_G = 0.3$) and the same strength of selection ($\Gamma = 2$), with different values of error (σ). **C)** The pattern of absolute fitness calculated following the same method used to process the experimental data. Parameter values match those used in Figure 4. **D)** Experimental estimates of absolute

fitness based on the patterns of stalk investment (Figure 4D) and the probability of fruiting body collapse (Figure 6A). The black points represent the estimates at the measured frequencies. The line represents the best fit line based on a cubic regression using these estimates, with the shaded region indicating one standard error on either side of the best fit line.

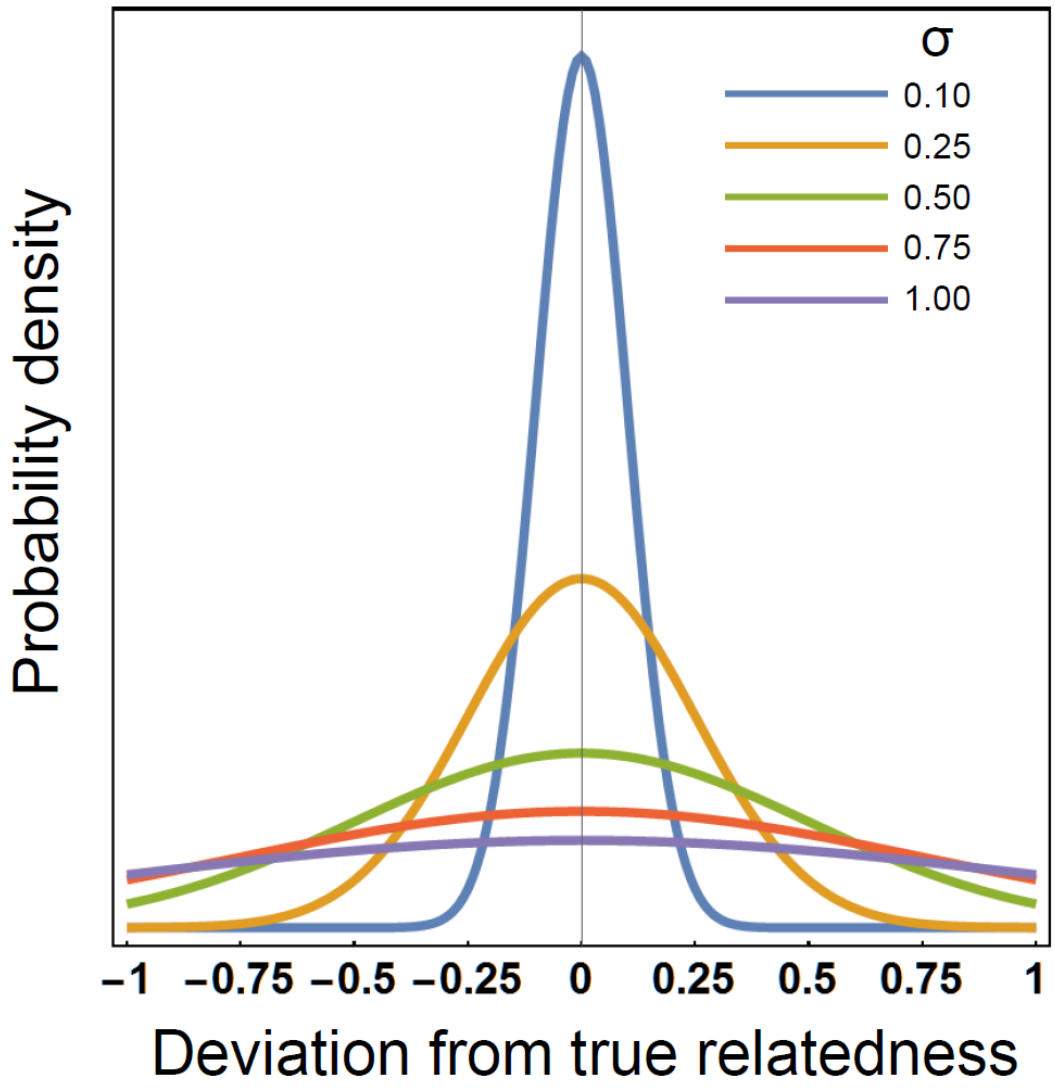


Figure S3. The distributions of errors in players' measurement of relatedness (frequency) that were used to generate patterns under imperfect information. Illustrated are five different probability density functions that differ in the level of error in the measurement of relatedness (frequency) (σ). The distributions are Gaussian with a mean of zero and a standard deviation given by σ. The deviations represent the difference between the relatedness (frequency) that player estimates for their group and their true relatedness. Because relatedness (frequency) is constrained to the range of zero to one, the distribution is necessarily truncated when used in calculations. For example, if a player's true relatedness is 0.25, the deviations will be truncated at −0.25 and 0.75 (where the mean of zero indicates that they have correctly measured their relatedness as 0.25).

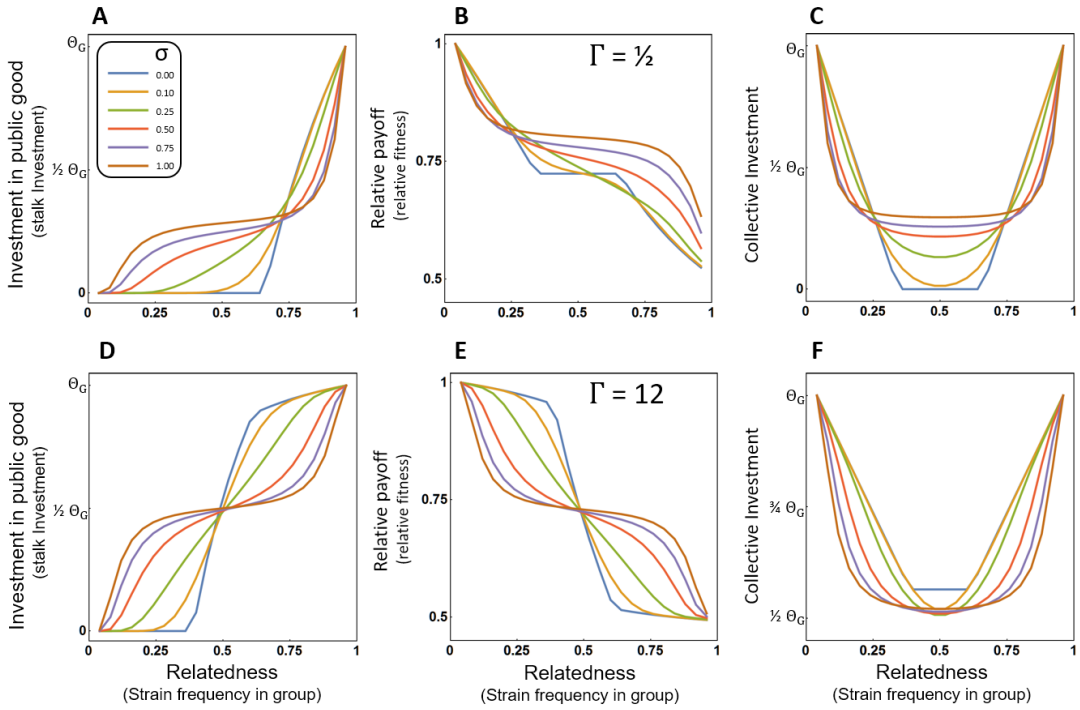


Figure S4. Predicted patterns of individual investment, relative payoffs and collective investment for different strengths of selection (Γ) and levels of error in measurement of relatedness (frequency) (σ). The general structure of the figures follows that of Figure S1, except the rows show the same relationships for two different strengths of selection (Γ = ½ for **A** to **C**, and Γ = 12 for **D** to **F**), which, combined with the values illustrated in Figure 3 essentially cover the entire range of parameter space. The lines in each figure show the pattern expected for varying amounts of error in the measurement of relatedness (frequency) (σ) (see the legend imbedded in the first figure). All examples were calculated for the same optimal level of collective investment (Θ_G = 0.3) since the exact optimum has a minor effect on the patterns (see Figure 2 and S1). The labels in parentheses relate the figures to the patterns expected for the *D. discoideum* system.

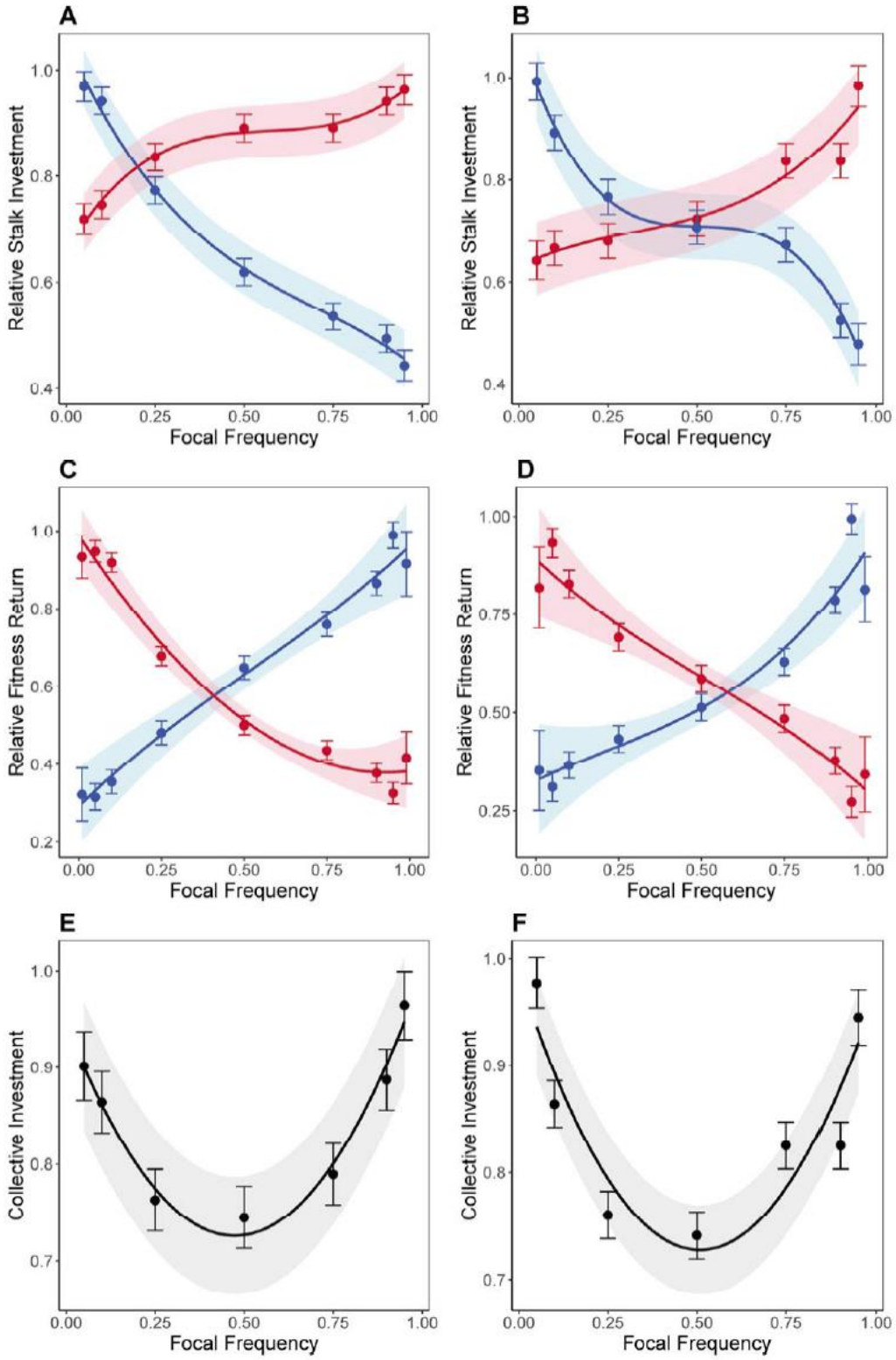


Figure S5. Patterns of stalk investment, relative fitness, and collective investment as a function of the frequency of a strain in a chimeric aggregation for two high resolution pairs of natural strains combined in chimeras. **A)** Stalk investment ($I_{i|p_i}$) by NC105.1 (in red) and

NC34.2 (in blue) in chimeric mixtures of the two, plotted as a function of the frequency of NC105.1 (designated as the focal strain) in the mix, **B)** Stalk investment ($I_{i|p_i}$) by NC63.2 (in red) and NC28.1 (in blue) in chimeric mixtures of the two, plotted as a function of the frequency of NC63.2 (designated as the focal strain) in the mix. **C)** Relative fitness (ρ_i) of NC105.1 (in red) and NC34.2 (in blue) in chimeric mixtures of the two, plotted as a function of the frequency of NC105.1 in the mix. **D)** Relative fitness (ρ_i) of NC63.2 (in red) and NC28.1 (in blue) in chimeric mixtures of the two, plotted as a function of the frequency of NC63.2 in the mix. **E)** Collective stalk investment (I_G) by chimeras composed of NC105.1 and NC34.2 as a function of the frequency of NC105.1 in the mix, **F)** Collective stalk investment (I_G) by chimeras composed of NC 63.2 and NC28.1 as a function of the frequency of NC63.2 in the mix. Each panel shows the estimated means (with their standard errors) at each frequency measured (with values estimated by the mixed model describe in the Methods, but using frequency as a categorical factor). In all cases, the y-axis values are scaled as proportions of the maximum value observed. In parts **A** to **D** the bold curve represents the best-fit estimated from the cubic regression model (here fitted to the estimated means) and the like-colored shaded regions give approximate 95%-confidence intervals around those curves. For parts **E** and **F** the curve represents the best-fit estimate from a quadratic regression model (fitted to the estimated means) and the shaded region gives the 95% confidence interval around that relationship.

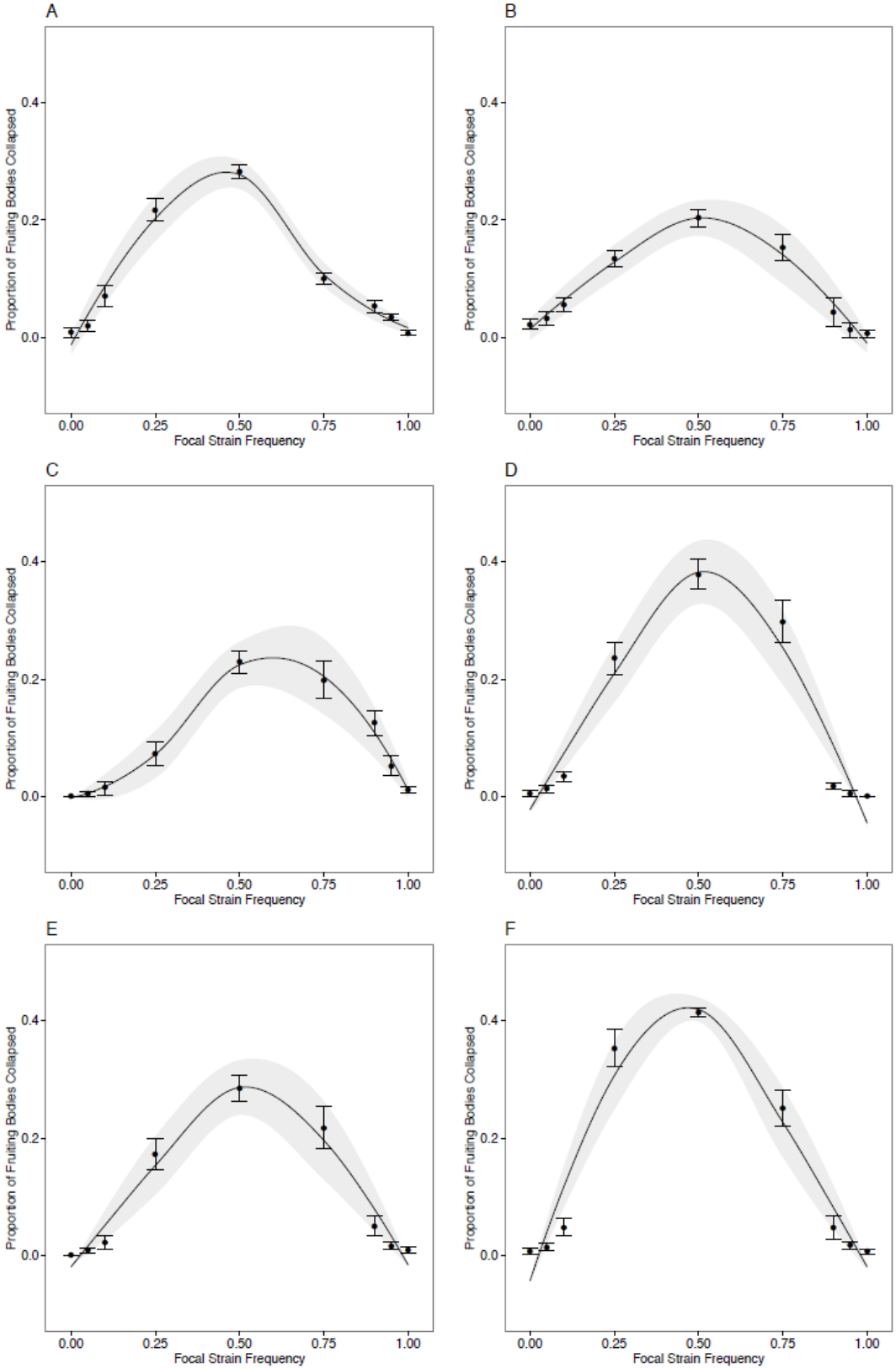


Figure S6. Spontaneous fruiting body collapse as a function of focal strain frequency for six strain pairs. Points represent the mean observations (with standard error bars) and the curve

illustrates the best-fit polynomial relationship (with 95%-confidence intervals as grey-shading). The six pairs appear as: **A)** NC28.1+NC105.1, **B)** NC60.1+NC99.1, **C)** NC34.2+NC105.1, **D)** NC99.1+NC105.1, **E)** NC60.1+NC34.2 and **F)** NC60.1+NC63.2.

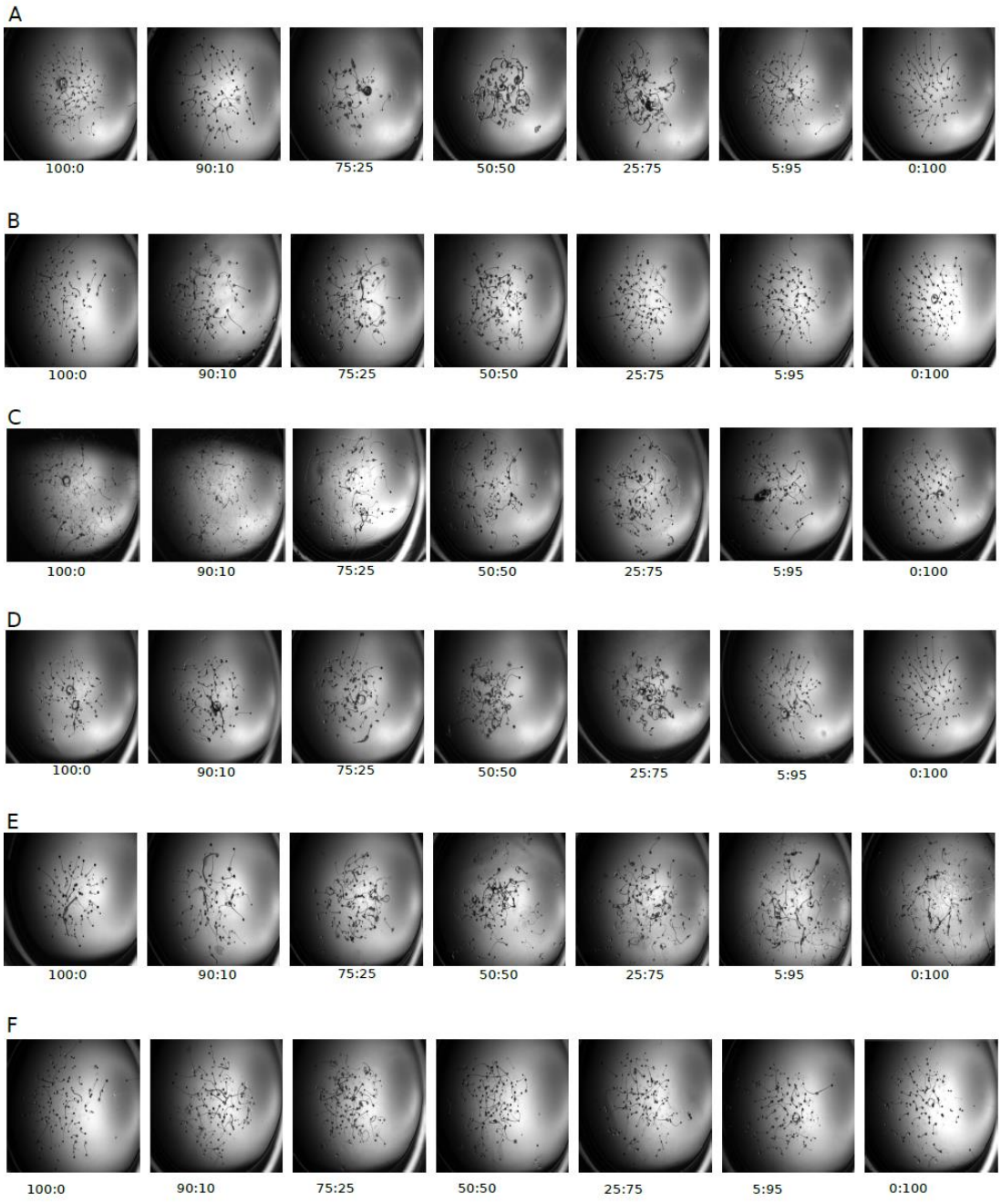


Figure S7. Representative images of fruiting bodies for six strain pairs across a range of focal strain frequencies. **A)** NC34.2+NC105.1, **B)** NC60.1+NC34.2, **C)** NC60.1+NC63.2, **D)** NC99.1+NC105.1, **E)** NC60.1+NC99.1 and **F)** NC28.1+NC105.1.

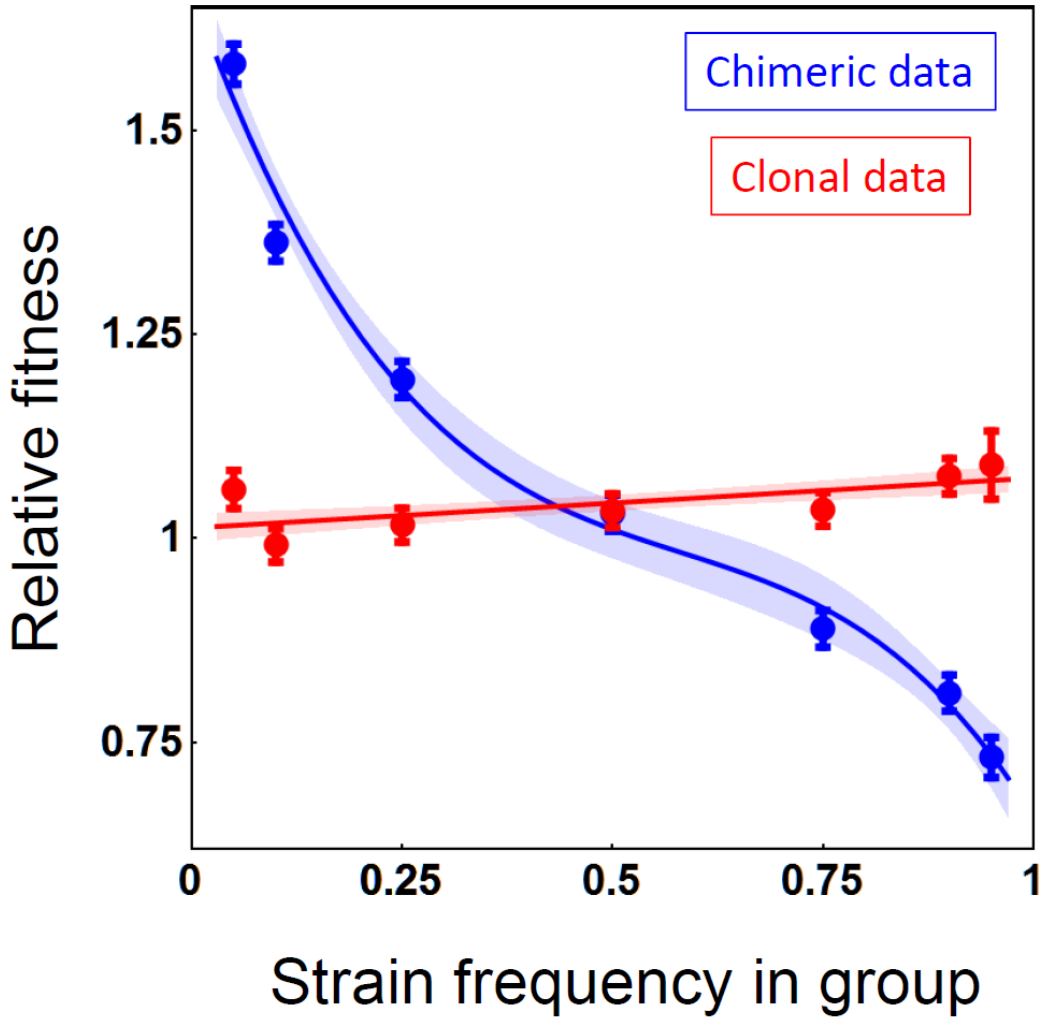


Figure S8. The pattern of relative fitness as a function of frequency in a group. The parts in blue correspond to the estimated pattern for chimeric mixes. They therefore match the pattern of relative fitness shown in Figure 4E and are included here for comparison. The parts in red correspond to the pattern for clonal self-mixes, which were estimated following the same method used to calculate the chimeric pattern, with the labelled cells considered as the 'focal' strain. The points indicate the means and the bars their standard errors, both estimated from a mixed model. The lines represent the best fit relationship (which is cubic for the chimeric data and linear for the clonal data), with the shaded region indicating one standard error on either side. The slope of the best fit line for the clonal data is not significant ($F_{1, 195} = 1.65$, $p = 0.2$, but is included as an illustration of the relationship.

Paper 2

In this supplement to the main text, we analyse the nature of error that strains have in estimating their relatedness to the group. In particular, we examine 1) consistent overestimation of relatedness by stains, and 2) bet-hedging strategies in response to error in estimating relatedness. The ESS for the error function used in the main-text is shown in Figure S1A for comparison to the different forms of error analysed here.

Overestimation of relatedness

In the main text we present an analysis of the game where players have imperfect information about their relatedness to the group. Here we analyse the game where players have a one-sided bias in their information, always over-estimating relatedness. This is achieved by modelling a population of players, each of which make a systematic overrepresentation of magnitude S_i of their relatedness to the group r_i. Relatedness is therefore defined as $r_i + S_i$, and truncated such that r_i is ≤ 1. This has the effect of making deviations from the true value for relatedness greatest when at intermediate relatedness (as any deviation is less likely to be truncated). In any one instance, each player plays the perfect information ESS for the relatedness it estimates.

We created a population of players, each with a value of S_i drawn from a biased distribution such that smaller biases were more likely than larger biases. For any value of r_i each strain has an ESS level of investment x_i. The average across a large population of strain was used as the population-wide ESS for systematic overestimation of relatedness.

The ESS for the average player with overestimation in relatedness is shown in Figure S1B. The obvious finding is that overestimation of relatedness leads to an increase in investment for any value of r_i, and investment is greater than 0 even when a player is very rare. Investment is also much more linear than the perfect information game, with no evidence

for the threshold switch between 0 investment and intermediate investment characteristic of the perfect information models.

Overestimation of relatedness is a plausible mechanism for the tragedy of the commons to be avoided, but seems unlikely to be relevant in *D. discoideum* based on several lines of evidence. Firstly, the signature of overestimation is that strains invest at the optimal level even when their true relatedness to the group is less than one, a fact that is inconsistent with empirical data (Madgwick *et al.*, 2018). A second defining feature of overestimation is that strains never reach a plateau of zero investment at low relatedness to the group, which is also inconsistent with the data. Further, whilst all errors in relatedness estimation for individuals are non-adaptive, consistent overestimations seems likely to be the easiest to 'fix' evolutionarily, as it could require a simple retuning of a gene that governs the response to information, rather than having to respond to substantial stochastic noise. For these reasons, we can reject a consistent overestimation of relatedness as being a relevant factor in the patterns of investment we observe in *D. discoideum*, and suggest that it's wider importance is likely to be reduced compared to other forms of error.

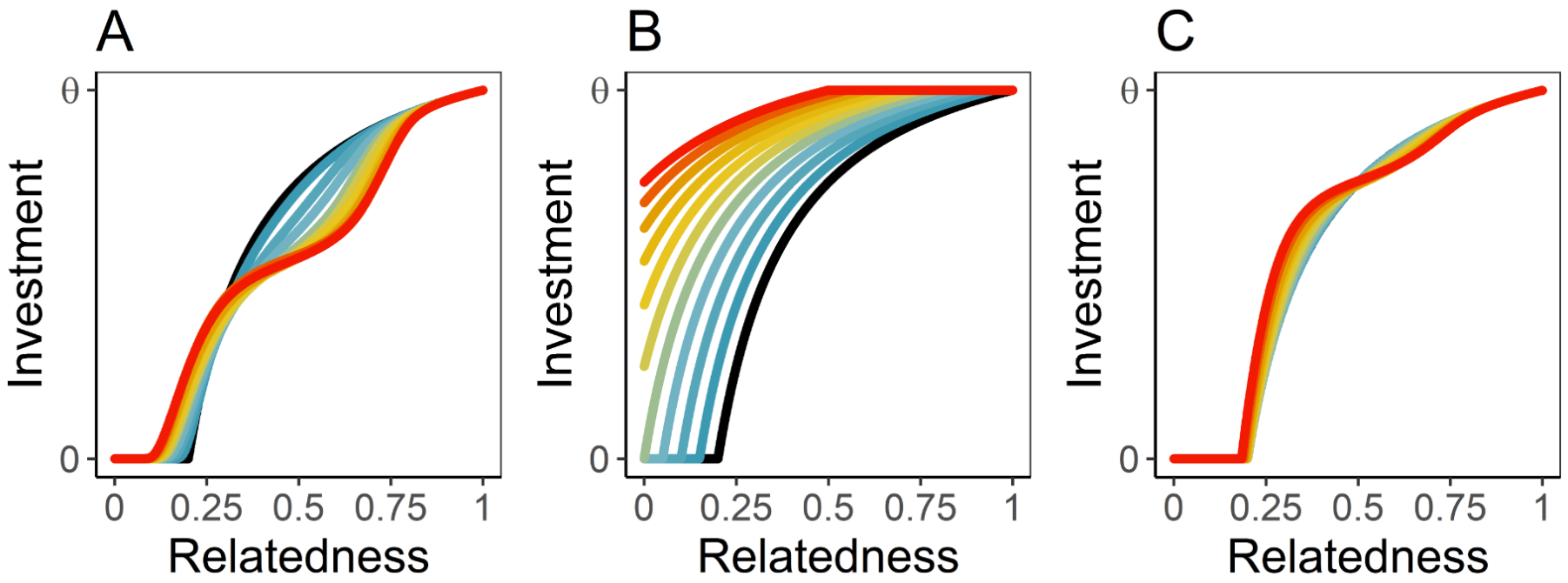

A
B
C
θ
0
Investment
0
0.25
0.5
0.75
1
Relatedness

Figure S1. Predictions for optimal investment with different kinds of imperfect information over a strains relatedness to the group. **A**) Noise: strains make stochastic, normally distributed errors in their estimation of relatedness to the group. **B**) Overestimation: strains consistently overestimate their relatedness to the group at all relatedness values. **C**) Bet-hedging: strains play the strategy that maximises their expected return given normally distributed probabilities of differences between their measured relatedness to the group and their true relatedness to the group. For all panels coloured lines represent variation in the magnitude of the errors, with blue representing small errors and red large errors.

Bet hedging

In biology bet-hedging is characterised through variation in the fitness of traits in different environments. If the environment is not known or predictable, then there is a trade-off between the fitness of different strategies in the different environments. The optimal strategy in this scenario can be bet-hedging – choosing a strategy that maximises fitness given the range of possible environments and their relative likelihood (Cohen, 1966). Bet-hedging can have important consequences on the evolution of cooperation (Kennedy *et al.*, 2018), and can be one of two main types of 1) maximise expected fitness across all environments, or 2) minimise variance in fitness (Philippi & Seger, 1989; Olofsson *et al.*, 2009).

In the main text, we present an analysis of imperfect information through players measuring their relatedness as a deviation from 'true' relatedness. For every value of 'true' relatedness, players used the average fixed ESS for 'measured' relatedness across a probability distribution of measured relatedness. As such, this is simply error in measurement, rather than a bet-hedging strategy.

We implement a bet-hedging strategy here by taking each value of *measured* relatedness r_i and creating a Gaussian probability distribution with a mean $\bar{x}$ equal to the player's measured relatedness to the group, and a standard deviation s that represents the error in estimating relatedness. Each value on this distribution therefore represents a possible 'true' relatedness that a player could have, with the probability distribution representing the relative likelihood of having each value of 'true' relatedness. We can then ask what strategy maximises expected fitness for any value of measured relatedness, given the possibility of true relatedness being at each different value. This is achieved for any value of measured relatedness r_i by finding the strategy $\hat{x}_i$ that has highest fitness on average given the probability distribution of possible 'true' relatedness. As such, we use a bet-hedging strategy that maximises expected fitness. A 'conservative' bet-hedging strategy that minimises the variance in fitness rather than expected fitness is not appropriate here, because the strategy that minimises the variance in fitness often has consistent extremely low fitness. The ESS solution is found using the 'brute-

force' iterative approach of calculating average fitness of each possible strategy, with Nelder-Mead optimisation to decrease computational time. As with imperfect information within the main text, we assume that players are best at estimating their relatedness when at the extremes of relatedness within the group.

The ESS for varying level of error in estimating relatedness is shown in Figure S1C. Lower errors are shown in blue, with the highest errors shown in red. The size of the error changes the range and relative likelihood of 'true' relatedness that a player could have for any measured relatedness. The general pattern is somewhat similar to stochastic errors in relatedness shown in the main text; players tend to overinvest compared to the perfect information model when rare and underinvest when common. Further, this effect is stronger when error is larger. The effect isn't however nearly as pronounced as stochastic error (Figure S1A). In this way, a bet hedging strategy gives only marginally different predictions to perfect information, which fit significantly worse to the data than the error model presented in the main text.

The players in the Collective Investment game are genetically distinct genotypes that invest a portion of their cells into the stalk (the public good). In practise, each individual cell becomes either a spore or a stalk cell. With information received from cell-cell contact, it is easy to see that any noise in the system will change the probability of each cell becoming either a spore or a stalk cell. In this way, the strategy is a function of the random noise in information experienced by each cell – which is the basic logic we used to model imperfect information in the main text. For bet-hedging to occur, the error that each cell experiences through variation in the probability of interacting with self would have to be centrally processed to produce an 'optimal' probability for a cell to become a stalk vs spore, something that seems unlikely to occur.

Frequency-dependant error in relatedness estimation

In the main text, we model imperfect information through a Gaussian error function based on the simple logic of the variance in probability of a cell meeting self across a range of frequencies leading to the greatest error in estimating relatedness when $\boldsymbol{r_i}$=0.5, with the magnitude of the error decreasing as $\boldsymbol{r_i}$ approaches either 0 or 1. However, given that there was no shape of the possible error function between these three points that should be prioritised *a priori*, we modelled a range of options, based on normal distributions that had been normalised to peak at 1. The range of shapes we modelled are shown in Figure S2.

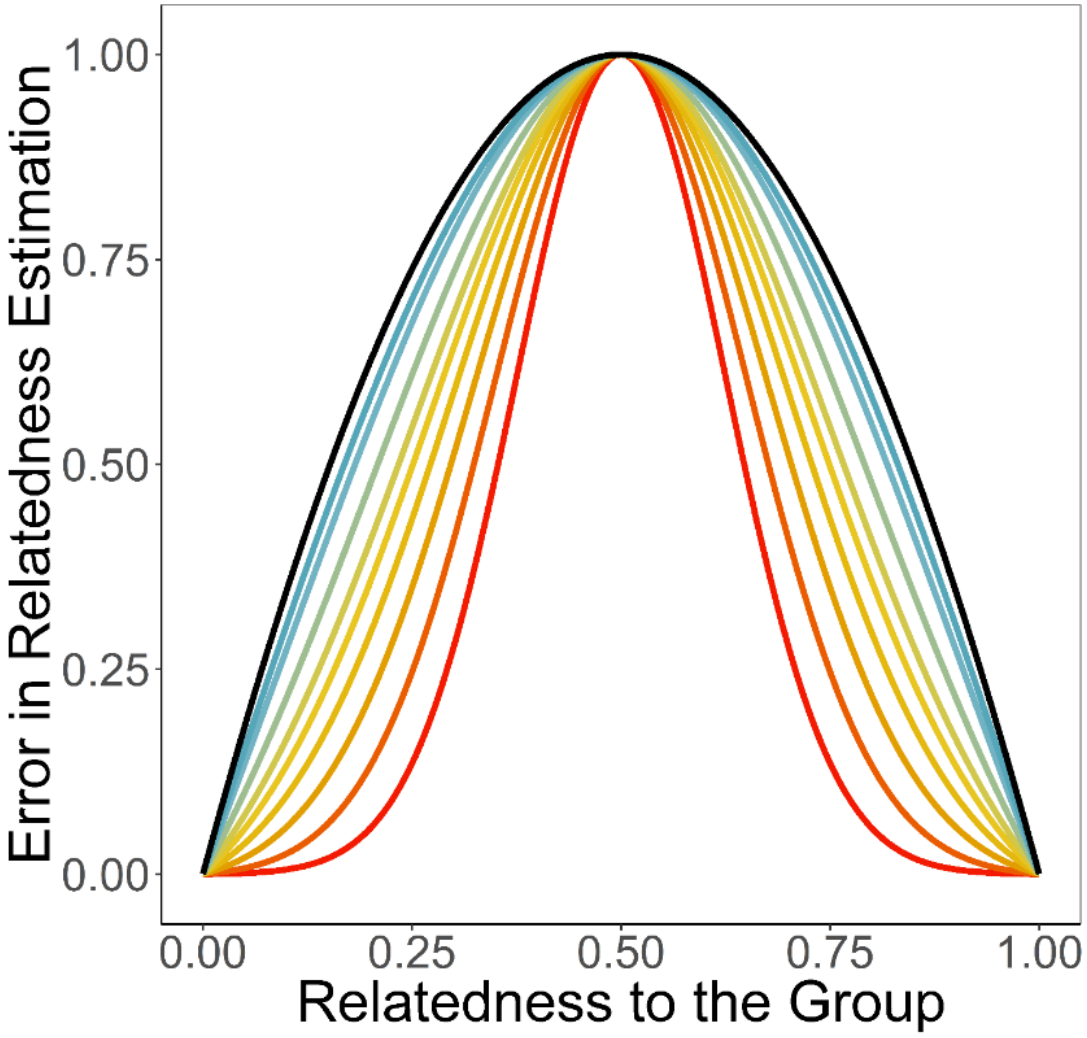


Figure S2. Distribution of frequency dependant noise. Strains are assumed to make frequency (whole-group relatedness) dependant errors in estimating their relatedness. Errors in identifying self vs non-self will always be greatest when one strain is at intermediate frequency. If interactions between types (self vs non-self) are random, error will scale with *p(1-p)* (black line). If interactions between types are non-random, the shape of the magnitude of error with frequency will change, modelled here by a normal deviation with increasing standard deviation as the lines move from blue to red.

References

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**: 119–129.

Kennedy, P., Higginson, A.D., Radford, A.N. & Sumner, S. 2018. Altruism in a volatile world. *Nature* **555**: 359–362. Nature Publishing Group.

Madgwick, P.G., Belcher, L.J., Stewart, B., Thompson, C.R.L. & Wolf, J.B. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. *Proc. Natl. Acad. Sci.* **115**: E4823–E4832.

Olofsson, H., Ripa, J. & Jonzén, N. 2009. Bet-hedging as an evolutionary game: The trade-off between egg size and number. *Proc. R. Soc. B Biol. Sci.* **276**: 2963–2969.

Philippi, T. & Seger, J. 1989. Hedging One's Evolutionary Bets, Revisited. *Trends Ecol. Evol.* **4**: 2–5.

Paper 3

Section 1: Solving the evolutionarily stable strategy

In the main-text the evolutionarily stable strategy (ESS) for an individual player was derived using a method that utilises the coevolutionarily stable strategy (coESS) for the group, which we analyse in further detail here. The coESS is an unfamiliar term (but see *e.g.* Eshel 1985) that we use to provide our novel analysis of a generalised public goods scenario. We need to use the coESS because of the difficulty of analysing an N player public goods game, where N could refer to any number of players, who could also have any distribution of frequencies. Each possible group has its own coevolutionary solution, which can be derived via a numerical analysis to solve the ESS (see end of section). However, we can solve the ESS using the coESS to derive an information open-solution because the coESS depends upon group-level variables that have known properties. (NB: It is possible to provide a closed-solution without need of the coESS when there are two players only, but this represents a special case that has been analysed elsewhere (Madgwick *et al.*, 2018).)

We provide a detailed solution to the ESS, using a specific notation with respect to a focal player (i) and a combined set of non-focal players(s) ($-i$). Within this notation, this means that the collective investment by the group would be written as the sum of both the focal player's investment and any investment arising from others: $x_G = p_i x_i + p_{-i} x_{-i}$. This notation is useful because it allows us to resolve the coESS analytically, and hence the ESS in turn. The ESS is solved by starting from the optimal strategy that maximizes fitness – *i.e.*, a player's 'best-response' ($\vec{x}_i$), which is dependent upon the level of investment made by other players. Players are assumed to have access to perfect information about their frequency in the group (p_i) and the investment from all other players ($p_{-i} x_{-i}$) and so, given that strategies are restricted to within the range $0 \leq \vec{x}_i \leq 1$, the best-response is solved by setting the derivative $d\omega_i / d\vec{x}_i = 0$ (where the collective investment can be written as $x_G = p_i \hat{x}_i + p_{-i} \hat{x}_{-i}$):

$$0 = bp_i^2 - cp_i - bcp_i p_{-i} x_{-i} - 2bcp_i^2 \vec{x}_i$$

$$\vec{x}_i = \frac{1}{2}\left[\frac{1}{c} - \frac{1}{p_i b} - \frac{p_{-i} x_{-i}}{p_i}\right] \quad \text{(S1.1)}$$

The best-response is the same as the evolutionarily stable strategy (ESS; $\hat{x}_i$) when all other players are also playing their best-response (*i.e.* other players are also playing the ESS: $\hat{x}_{-i}$). To solve the ESS, the expression in equation (S1.1) must be resolved simultaneously for all players, which collectively behave as per the coevolutionarily stable strategy (coESS; $\hat{x}_G$) for the group. Because the ESS describes the full conditional strategy as frequency (p_i) varies, the ESS must be solved simultaneously across frequencies as well as across different players. When the number of players is not specified, we do not believe that it is possible to solve the investment by a given player across all frequency combinations, which is supported by the logic that each player/frequency set has its own coevolved solution. Instead we have to use a novel method to discern the critical variables underlying how a player's strategy changes in response to their own frequency and the number/frequencies of other players.

At the equilibrium where all players display the conditional ESS, not all players necessarily contribute toward the coESS because some players have too low a frequency to benefit from investment into public goods. Players in any given social context can be divided into two groups at the coESS: a set of n 'investors' that contribute toward collective investment and a set of $N - n$ non-investors that do not contribute (where N is the number of players in the game). Formally, given that the best-response (eqn. S1.1) can be negative (even if an individual cannot actually display negative investment because this would be nonsensical), the set of n investors includes all players that have a non-negative investment in the ESS ($\hat{x}_i \geq 0$), which includes the non-investing strategy where $\hat{x}_i = 0$. Therefore, players can be divided into two classes based on Equation S1.1: 'investors' that have investment

values $\hat{x}_i \geq 0$ and non-investors that have an optimal investment below zero $\hat{x}_i < 0$, but which is constrained to be zero. However, for simplicity and given that the frequency where $\hat{x}_i = 0$ exactly is infinitesimally small, we can more generally describe players as one of n investors ($\hat{x}_i > 0$) or $N - n$ non-investors (whose real strategy is constrained to $\hat{x}_i = 0$) without worrying about this distinction (as we do in the main text) – and we use the terms investors and non-investors going forward.

By summing up the frequency of all investors, the collective frequencies of investors is $p_n = \sum_{j=1}^{n} p_j$ whilst the collective frequencies of non-investors is $1 - p_n$. Thus, the coESS can be calculated as the sum of the contributions by ESS investors:

$$\begin{aligned}\hat{x}_G &= \sum_{i=1}^{n} p_i \hat{x}_i \\ &= \sum_{i=1}^{n} \frac{1}{2}\left[\frac{p_i}{c} - \frac{1}{b} - p_{-i}\hat{x}_{-i}\right] \\ &= \frac{1}{2}\left[\frac{p_n}{c} - \frac{n}{b}\right] - \sum_{i=1}^{n} \frac{p_{-i}\hat{x}_{-i}}{2}\end{aligned} \tag{S1.2}$$

Given that $\hat{x}_G = p_i\hat{x}_i + p_{-i}\hat{x}_{-i}$ by definition, the act of summing $p_{-i}\hat{x}_{-i}$ with respect to the i^{th} focal player n times is the same as $\sum_{i=1}^{n} \hat{x}_G - p_i\hat{x}_i = n\hat{x}_G - \sum_{i=1}^{n} p_i\hat{x}_i = n\hat{x}_G - \hat{x}_G$, so this leads to:

$$\hat{x}_G = \frac{1}{2}\left[\frac{p_n}{c} - \frac{n}{b} - \hat{x}_G(n-1)\right] \tag{S1.3}$$

$$= \left[\frac{1}{n+1}\right]\left[\frac{p_n}{c} - \frac{n}{b}\right]$$

Equation (S1.3) corresponds to equation (2) in the main-text. As the coESS reflects ESS behaviour for a group of investing players, we can substitute the coESS into the best-response (eqn. S1.1), using $p_{-i}\hat{x}_{-i} = \hat{x}_G - p_i\hat{x}_i$, to resolve the ESS:

$$\hat{x}_i = \frac{1}{2}\left[\frac{1}{c} - \frac{1}{p_i b} - \frac{\hat{x}_G - p_i\hat{x}_i}{p_i}\right]$$

$$= \frac{1}{c} - \frac{1}{p_i b} - \frac{\hat{x}_G}{p_i}$$

$$= \frac{1}{c} - \frac{1}{p_i b} - \left[\frac{1}{p_i(n+1)}\right]\left[\frac{p_n}{c} - \frac{n}{b}\right] \quad \text{(S1.4)}$$

Equation (S1.4) corresponds to equation (2) in the main-text. Given that players' investments are constrained to be non-negative ($\hat{x}_i \geq 0$), these solutions can only calculate the quantity of investment at the ESS if players can be classified as being one of the n investors or $N - n$ non-investors. A player's strategy can be categorised based on their frequency by solving the boundary condition where individuals would switch from being a non-investor to an investor, which corresponds to the value where $\hat{x}_i = 0$ (using eqn. S1.3). Given that a player is an investor when $\hat{x}_i = 0$, a player can be classified as an investor based on their frequency when:

$$p_i \leq \frac{p_n b + c}{b(n+1)} \quad \text{(S1.5)}$$

Equation (S1.5) corresponds to an equation found in Table 2 of the main-text. Because this

limit is necessarily self-referential to the collective frequency and number of investors (p_n, n), the ESS (eqn. S1.4) represents a closed-solution only insofar as players are correctly classified as investors or non-investors. Because the ESS has this self-referential property, solving for the exact quantity of investment at the ESS when there is more than one investing player requires a numerical solution given a particular distribution of frequency across a definite number of players.

The additional insight into the ESS gained from comparison of Equation (S1.1) to the form of Equation (S1.4) permits a simple algorithm to be used to categorise players, which facilitates the search for a numerical solution to the ESS because it is computationally efficient. This simple algorithm relies upon a single assumption, which is known to be true: a player with a higher frequency invests more or equally to that of a player with lower frequency. The algorithm can solve the quantity of ESS investment of each player by ranking players in terms of their frequency (high to low) and 'testing' each player to examine whether or not they have a non-negative investment strategy in the new social context (*i.e.* when that player is included in the calculation of the collective frequency and number of investors; p_n, n). If the player has a non-negative investment strategy then they are an investor and the next highest frequency player can then be tested. The algorithm can stop searching for new players to add into the investor class once the next highest frequency player fails to have a positive investment. At this point, the candidate player is not added into the investor class and the ESS is fully resolved with the previous calculations of the collective frequencies and number of investors prior to the current test. To demonstrate that this algorithm is indeed effective, an alternative method using an unbiased iterator to solve the ESS can be utilised.

Section 2: Deriving a marginal Hamilton's rule

Although we point to numerous similarities between inequalities utilising the ESS and Hamilton's rule, their formal relationship is more complicated. Under quantitative strategies, we can use Hamilton's rule to solve the ESS, but the rule must be framed in a marginal form (Taylor & Frank, 1996; Frank, 2013) which, given relatedness (r), describes the ESS as the equalisation of the specially-constructed marginal costs (C_m) and marginal benefits (B_m) of cooperation: $rB_m - C_m = 0$. We say 'specially-constructed' to emphasise that these marginal costs and benefits (C_m, B_m) are not the same as the cost and benefit terms (c, b) in our model, and so we have to calculate them in the terms of our existing model. The two differ because the cost and benefit terms in our model are not defined with respect to fitness, but rather, are parameters that interact (multiplicatively) to determine fitness, whereas the cost and benefit terms that are relevant to Hamilton's rule are defined in terms of additive components of fitness.

In-keeping with our model, we can rearrange the ESS (eqn. 5), which has been solved by taking the derivative of fitness with respect to strategy derivative ($d\omega_i/dx_i$) given an equilibrium quantity of collective investment by the group ($\hat{x}_G$), into the marginal form:

$$p_i b - c - bc(p_i \hat{x}_i + \hat{x}_G) = 0 \qquad \text{(S2.1)}$$

On the right-hand-side, the first two parts, $p_i b - c$, are equivalent to the classic form of Hamilton's rule (Charnov, 1977), where a non-zero investment strategy is positively selected if the product of the public benefit from investment (b) and frequency/relatedness (p_i) outweighs the private cost (c). Yet it should be remembered that p_i is not a kinship coefficient of relatedness, but instead is 'whole-group relatedness' (Hamilton, 1975; Pepper, 2000), ranging between zero and one. The third part of this marginal Hamilton's rule, $-bc(p_i \hat{x}_i + \hat{x}_G)$, is equivalent to other derivations of Hamilton's rules that consider the synergistic impact

of the social context on an individual (Queller, 1994; Frank, 1998), which, in our model, reflects how a focal player's motivation to invest into public goods declines both in response to increasing their own level of investment ($p_i\hat{x}_i$) and the investment afforded all players within the group including themselves ($\hat{x}_G$). From a single player's perspective, this means that increasing their own investment into public goods by some marginal unit of resources has a two-fold effect on disincentivising any further investment by that player, which stems from that player shifting the unit of resource both toward the production of public goods (and hence incentivising them to exploit public goods more) and away from the utilization of public goods (and hence decreasing their receipt of the benefits from public goods and thereby incentivising them to exploit public goods more again).

In the marginal Hamilton's rule, the social context of investment provided by others is a property of both the marginal costs (C_m) and marginal benefits (B_m) of cooperation. Thus, ignoring the connection between personal ($\hat{x}_i$) and collective ($\hat{x}_G$) investment (see Supplement 1), the marginal form of Hamilton's rule can be written by grouping the relatedness-weighted (*i.e.* multiplied by p_i) benefits and the relatedness-independent costs:

$$p_i b(1-\hat{x}_i) = c(1+b\hat{x}_G) \tag{S2.2}$$

which implies that $B_m = b(1-\hat{x}_i)$ and $C_m = c(1+b\hat{x}_G)$. By including additional factors beyond the direct costs and benefits of strategic behaviour, the marginal Hamilton's rule can be used to derive a solution to the ESS (Grafen, 1985; Queller, 1992; Taylor & Frank, 1996; Frank, 2013). However, because of the connection between personal ($\hat{x}_i$) and collective ($\hat{x}_G$) investment, this method arrives at the same circularity as in Supplement 1, requiring the adoption of novel method to derive a more informative ESS and needing numerical simulation.

References

Charnov, E.L. 1977. An elementary treatment of the genetical theory of kin-selection. *J. Theor. Biol.* **66**: 541–550.

Eshel, I. 1985. Evolutionary Genetic Stability of Mendelian Segregation and the Role of Free Recombination in the Chromosomal System. *Am. Nat.* **125**:412–420.

Frank, S. A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**:520–522.

Frank, S.A. 1996. Policing and group cohesion when resources vary. *Anim. Behav.* **52**: 1163–1169.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press.

Frank, S.A. 2013. Natural selection. VII. History and interpretation of kin selection theory. *J. Evol. Biol.* **26**:1151–1184.

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surv. Evol. Biol.* **2**:28–89.

Hamilton, W.D. 1975. Innate Social Aptitudes of Man: an approach from evolutionary genetics. In: *ASA Studies 4: Biosocial Anthropology* (R. Fox, ed), pp. 133–153. Malaby Press.

Madgwick, P. G., Belcher, L.J., Stewart, B., Thompson, C.R.L. & Wolf, J.B. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. *Proc. Natl. Acad. Sci.***115**:E4823–E4832.

Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.

Queller, D.C. 1992. Quantitative Genetics, Inclusive Fitness, and Group Selection. Am. Nat. **139**:540–558.

Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**:70–73.

Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**:27–37.

Paper 5

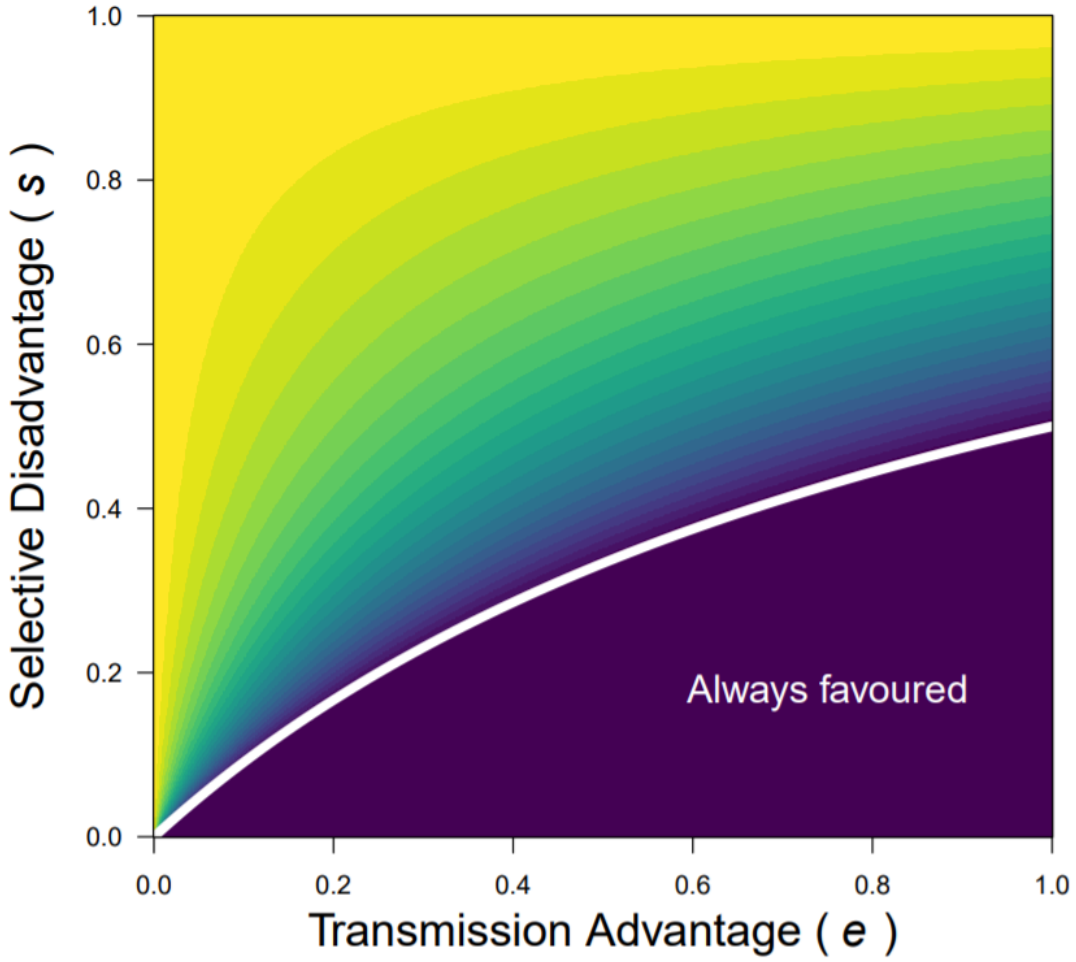


Figure S1. The frequency threshold above which a drive allele has higher fitness than a nondrive allele (see eqn. 2) across parameter space (of the selective disadvantage, $-s$, and the transmission advantage, e) in the single locus scenario. The frequency threshold is colour-scaled between <0 (purple) to 1 (yellow). The white line represents the parameter values where the drive and nondrive alleles have equal fitness: $s = e/(1 + e)$. When the inequality (eqn. 2) is negative (shaded purple), the drive allele always has higher fitness than the nondrive allele ($\omega_D > \omega_d$; hence the area below the white line is labelled 'Always favoured'). The progression of the colour scale from purple to yellow indicates the frequency above which the drive allele has higher fitness than the nondrive allele (from low to high).

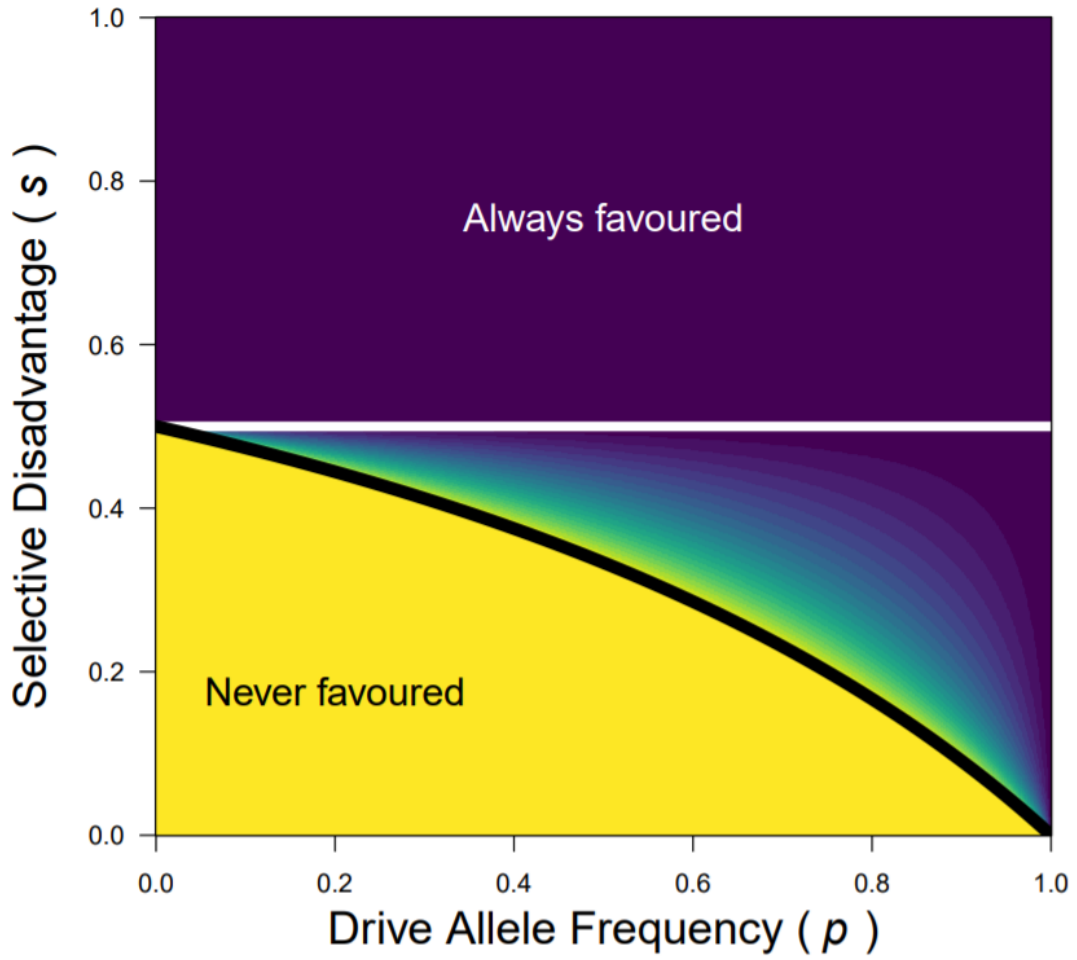
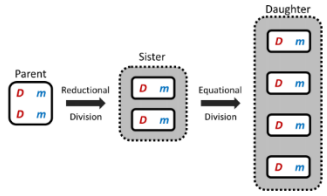
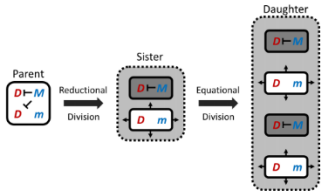
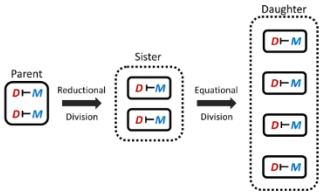
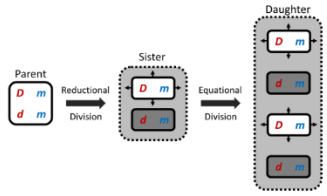
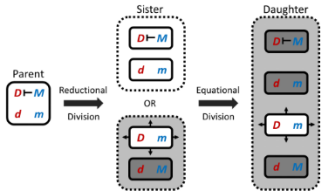
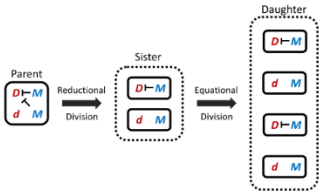
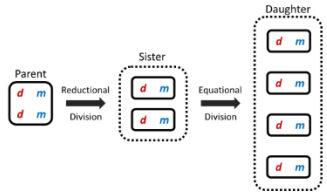
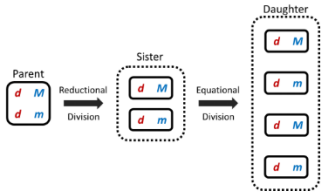
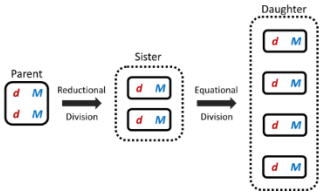


Figure S2. The frequency threshold above which a modifier allele that partially suppresses a drive allele by blocking toxin production has higher fitness than a nonmodifier allele (see eqn. 10) across parameter space (of the selective disadvantage, $-s$, and the drive allele's frequency, f_D, when the transmission advantage is maximal, $e = 1$). The frequency threshold is colour-scaled between <0 (purple) to 1 (yellow) to indicate the frequency above which the modifier allele has higher fitness than the nonmodifier allele (from low to high). When the inequality (eqn. 10) is negative (purple), the modifier allele always has higher fitness than the nonmodifier allele ($\omega_M > \omega_m$), which occurs when $s > e/(1+e)$ or, as $e = 1$, $s >$ ½ (hence the area above the horizontal white line is labelled 'Always favoured'). When the inequality (eqn. 10) is greater than 1, the modifier allele never has higher fitness than the nonmodifier allele ($\omega_M < \omega_m$), which occurs when $s > e(1-f_D)/(1+e(1-f_D))$ or, as $e = 1$, $s > (1-f_D)/(2-f_D)$ (hence the area above the black line is labelled 'Never favoured'). In

between the two lines the progression of the colour scale from purple to yellow indicates the frequency above which the modifier allele has higher fitness than the nonmodifier allele (from low to high)

A)	mm	Mm	MM
DD			
Dd			
dd			

1

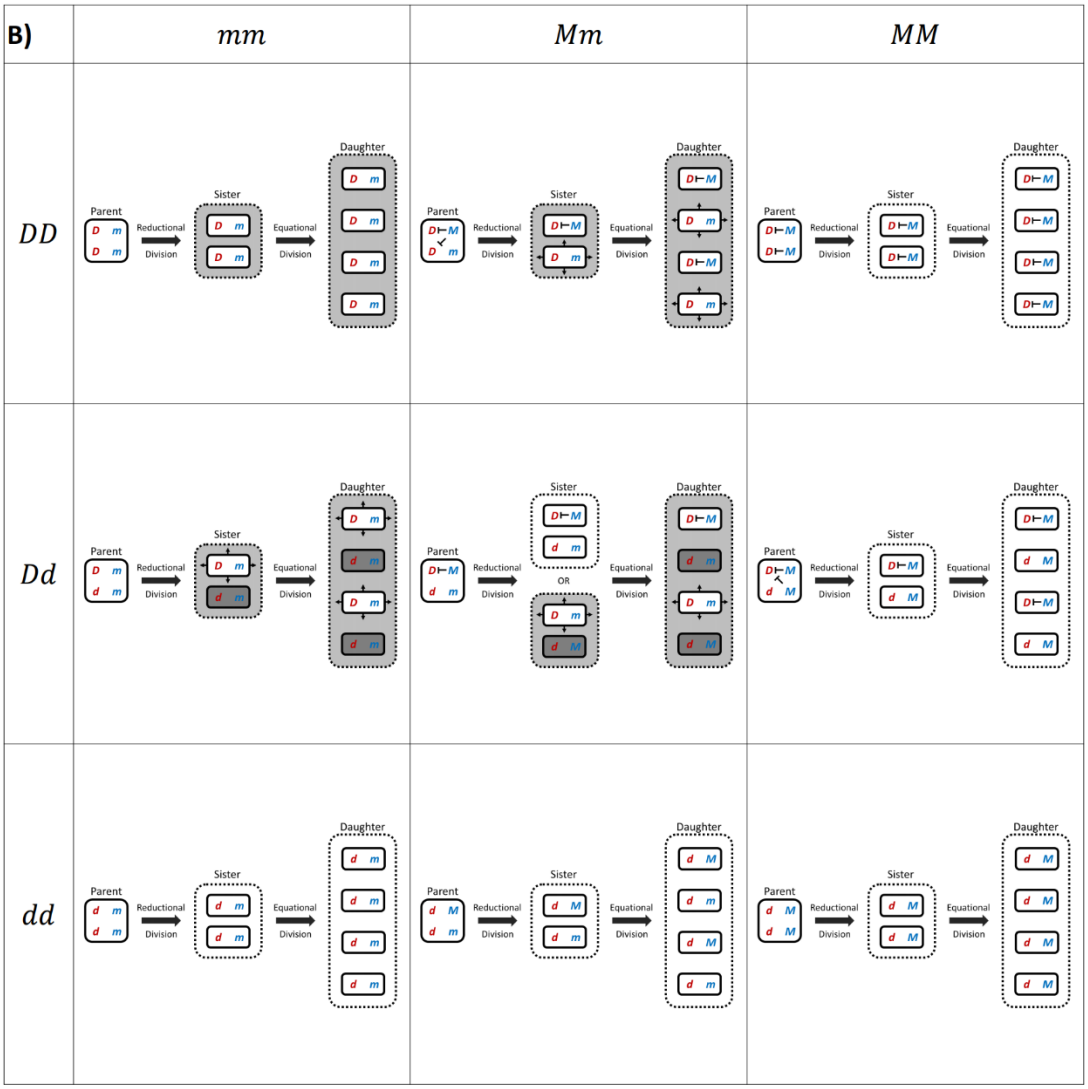


Figure S3. The potential for different types of game-killing gene drive throughout the steps of the meiotic divisions across genotypic combinations (indicated in a tabular form with respect to the drive and modifier loci). During meiosis, a diploid parental cell undergoes a first reductional division into two haploid sister cells followed by a second (mitosis-like) equational division into four haploid daughter cells. Different cells are presented with solid outlines, and different interactions between cells are presented with dashed outlines (except for the possible interaction between the parent cell and the sister cells it produced). Alleles at the drive locus are presented in red (D = drive and d = nondrive) and alleles at the modifier locus are presented in blue (M = modifier and m = nonmodifier). A drive allele for a toxin-antitoxin pair can distort its transmission to the next generation by producing a toxin (indicated

by out-facing arrows from the producer cell) that leads to the presence of the toxin within that given interaction (indicated by the light-grey shading in dashed outlined box). The presence of the toxin kills gametes that do not also produce the antixotin with cell death indicated by the dark-grey shading in solid outlined boxes). The modifier allele (M) only has an effect on the drive (D) allele if it is within the same cell, which is indicated by the turnstile symbol $\vdash$ that indicates that the drive allele has blocked some function of the drive allele (with different scenarios illustrated in A and B). As such, in the double heterozygote scenario presented here, there is no potential for pre-meiotic killing, where the diploid parent cell produces the toxin that kills haploid sister or daughter cells, because the modifier blocks toxin production. However, there is potential for meso-meiotic killing between sister cells (of which there are two potential genotypic combinations represented by different interactions in dashed boxes) and post-meiotic killing between daughter cells (ignoring the potential for daughter cells to kill sister cells in this simplified diagram). **A)** The case of a fully suppressing modifier that blocks toxin and antitoxin production by the drive allele, rendering it susceptible to the toxin produced by other cells. **B)** The scenario of a partially suppressing modifier blocks that toxin production only, and so remains resistant to the toxin produced by other cells.

Table S1. Fitness of two-locus genotypes under pre-meiotic killing, where the drive allele produces the toxin in the diploid cell (either in parental or somatic cells) making the modifier allele dominant to the nonmodifier allele (which is equivalent with modifiers for both full and partial suppression). The 'Genotype' column lists the nine possible unordered genotypes, whilst the 'gametes' column refers to the haploid gamete types that that genotype produces. The 'Fitness' column gives the relative production of the different gametes contributed to the next generation by that genotype. There are two loci (drive and modifier) which have two alleles: the drive locus has a drive allele D, which encodes a toxin and antitoxin, and nondrive allele d, while the modifier locus has a modifier allele M, which blocks the production of both the toxin and antitoxin, and a nonmodifier allele m. Within an interaction, when the toxin is produced it leads to a selective disadvantage $(-s)$ to all gametes and a transmission advantage $(+e)$ for those gametes that also produce the antitoxin, alongside a corresponding transmission disadvantage $(-e)$ for those gametes that do not produce the antitoxin. Shading is used to separate the sets of gametes produced by each diploid genotype.

Genotype	**Gametes**	**Fitness**
$DDMM$	DM	1
$DDMm$	DM	1
	Dm	1
$DDmm$	Dm	$1-s$
$DdMM$	DM	1
	dM	1
$DdMm$	DM	1
	Dm	1
	dM	1
	dm	1
$Ddmm$	Dm	$(1+e)(1-s)$
	dm	$(1-e)(1-s)$
$ddMM$	dM	1
$ddMm$	dM	1
	dm	1
$ddmm$	dm	1

Table S2. Fitness of two-locus genotypes under post-meiotic killing, where the drive allele produces the toxin in the haploid daughter cells (*i.e.* the gametes) making the modifier allele (for either full or partial suppression) recessive to the nonmodifier allele. The 'Genotype' column lists the nine possible unordered genotypes, whilst the 'gametes' column refers to the haploid gamete types that that genotype produces. The 'Fitness' column gives the relative production of the different gametes contributed to the next generation by that genotype. There are two loci (drive and modifier) which have two alleles: the drive locus has a drive allele D, which encodes a toxin and antitoxin, and nondrive allele d, while the modifier locus has a modifier allele M and a nonmodifier allele m. Within an interaction, when the toxin is produced it leads to a selective disadvantage $(-s)$ to all gametes and a transmission advantage $(+e)$ for those gametes that also produce the antitoxin, alongside a corresponding transmission disadvantage $(-e)$ for those gametes that do not produce the antitoxin. Shading is used to separate the sets of gametes produced by each diploid genotype. **A)** The modifier allele, M, causes full suppression of the drive locus by blocking the production of the toxin and the antitoxin. **B)** The modifier allele, M, causes partial suppression of the drive locus by blocking the production of the toxin but not the antitoxin.

A)

Genotype	**Gametes**	**Fitness**
$DDMM$	DM	1
$DDMm$	DM	$(1-e)(1-s)$
	Dm	$(1+e)(1-s)$
$DDmm$	Dm	$1-s$
$DdMM$	DM	1
	dM	1
$DdMm$	DM	$(1-e)(1-s)$
	Dm	$(1+3e)(1-s)$
	dM	$(1-e)(1-s)$
	dm	$(1-e)(1-s)$
$Ddmm$	Dm	$(1+e)(1-s)$

	dm	$(1-e)(1-s)$
$ddMM$	dM	1
$ddMm$	dM	1
	dm	1
$ddmm$	dm	1

B)

Genotype	**Gametes**	**Fitness**
$DDMM$	DM	1
$DDMm$	DM	$1-s$
	Dm	$1-s$
$DDmm$	Dm	$1-s$
$DdMM$	DM	1
	dM	1
$DdMm$	DM	$(1+e)(1-s)$
	Dm	$(1+e)(1-s)$
	dM	$(1-e)(1-s)$
	dm	$(1-e)(1-s)$
$Ddmm$	Dm	$(1+e)(1-s)$
	dm	$(1-e)(1-s)$
$ddMM$	dM	1
$ddMm$	dM	1
	dm	1
$ddmm$	dm	1

Table S3. Fitness of two-locus genotypes under meso-meiotic killing, where the drive allele produces the toxin in the haploid sister cells (for either full or partial suppression). The 'Genotype' column lists the nine possible unordered genotypes, whilst the 'gametes' column refers to the haploid gamete types that that genotype produces. The 'Fitness' column gives the relative production of the different gametes contributed to the next generation by that genotype. There are two loci (drive and modifier) which have two alleles: the drive locus has a drive allele D, which encodes a toxin and antitoxin, and nondrive allele d, while the modifier locus has a modifier allele M, which blocks the production of both the toxin and antitoxin, and a nonmodifier allele m. Within an interaction, when the toxin is produced it leads to a selective disadvantage $(-s)$ to all gametes and a transmission advantage $(+e)$ for those gametes that produce the antitoxin, alongside a corresponding transmission disadvantage $(-e)$ for those gametes that do not produce the antitoxin. Shading is used to separate the sets of gametes produced by each diploid genotype. **A)** The modifier allele, M, causes full suppression of the drive locus by blocking the production of the toxin and the antitoxin. **B)** The modifier allele, M, causes partial suppression of the drive locus by blocking the production of the toxin but not the antitoxin.

A)

Genotype	**Gametes**	**Fitness**
$DDMM$	DM	1
$DDMm$	DM	$(1-e)(1-s)$
	Dm	$(1+e)(1-s)$
$DDmm$	Dm	$1-s$
$DdMM$	DM	1
	dM	1
$DdMm$	DM	1
	Dm	$(1+e)(1-s)$
	dM	$(1-e)(1-s)$
	dm	1
$Ddmm$	Dm	$(1+e)(1-s)$

	dm	$(1-e)(1-s)$
$ddMM$	dM	1
$ddMm$	dM	1
	dm	1
$ddmm$	dm	1

B)

Genotype	**Gametes**	**Fitness**
$DDMM$	DM	1
$DDMm$	DM	$1-s$
	Dm	$1-s$
$DDmm$	Dm	$1-s$
$DdMM$	DM	1
	dM	1
$DdMm$	DM	1
	Dm	$(1+e)(1-s)$
	dM	$(1-e)(1-s)$
	dm	1
$Ddmm$	Dm	$(1+e)(1-s)$
	dm	$(1-e)(1-s)$
$ddMM$	dM	1
$ddMm$	dM	1
	dm	1
$ddmm$	dm	1

Section 1: Analysis of linkage disequilibrium

The main-text undertakes an analysis of the conditions favouring a drive allele and disfavouring a modifier that either fully or partially suppresses the drive allele under the simplifying assumption of linkage equilibrium. It is reasonable to expect linkage disequilibrium between alleles at the drive and modifier loci, which could alter the effectiveness of the gene drive by gamete-killing. Here, we undertake a simple analysis of linkage disequilibrium and provide a case for support of our analysis in the main-text.

To analyse how natural selection builds linkage disequilibrium we need to consider haplotype frequencies. In a system with two loci each with two alleles (as here), linkage disequilibrium (D) is the deviation between the observed haplotype frequencies and those expected based on random combinations of alleles (Lewontin, 1974; Hartl & Clark, 2006):

$$f_{DM} = f_D f_M + D \quad \text{(S1.1a)}$$

$$f_{Dm} = f_D f_m - D \quad \text{(S1.1b)}$$

$$f_{dM} = f_d f_M - D \quad \text{(S1.1c)}$$

$$f_{dm} = f_d f_m + D \quad \text{(S1.1d)}$$

It, therefore, follows that linkage disequilibrium is calculated:

$$D = f_{DM} f_{dm} - f_{Dm} f_{dM} \quad \text{(S1.2)}$$

Linkage disequilibrium can be built by natural selection due to additive and epistatic effects (Lewontin, 1974; Hastings, 1985; Hallgrímsdóttir & Yuster, 2008). Linkage disequilibrium

can arise from the additive effect because of the simultaneous selection of an allele at each locus, which creates interference between the selection at each locus. We anticipate that this is likely to occur in many scenarios of gamete-killing gene drive where both the drive and modifier alleles are favoured at their respective loci. However, this form of linkage disequilibrium is uninteresting because it does not favour an association between alleles due to the formation of a co-adapted haplotype – and so it does not change the outcome of natural selection (rather, it simply slows the rate of evolution). Linkage disequilibrium can also arise from the epistatic effect because some combinations of alleles form a co-adapted haplotype that has higher fitness than would be expected due to the additive contributions of each allele. The epistatic effect (*E*) is defined as (Hastings, 1985; Nagylaki, 1993):

$$E = (\omega_{DM} + \omega_{dm}) - (\omega_{Dm} + \omega_{Dm}) \quad \text{(S1.3)}$$

From this standpoint, the expected relationship between linkage disequilibrium (after one generation of selection from linkage equilibrium) and epistasis is:

$$D = f_D f_d f_M f_m E/\bar{\omega}^2 \quad \text{(S1.4)}$$

Thus, by calculating D and extracting E in different scenarios, we can assess the extent to which natural selection builds linkage disequilibrium because of co-adapted haplotypes. Logically, considering the five scenarios of gamete-killing gene drive, it is the epistatic effect that is critical to describing whether or not linkage disequilibrium would change the overall result because the modifier allele has the power to change the fitness effect of the drive allele (*i.e.* there is epistasis), which could potentially favour an association between alleles that changes the outcome of natural selection by making some haplotypes more common.

Table S1.1. The fitness of the four haplotypes of the combinations of drive/nondrive and modifier/nonmodifier alleles (ω_{DM}, ω_{Dm}, ω_{dM}, ω_{dm}) under different gamete-killing scenarios (see Tables S1, S2 and S3). **A)** Pre-meiotic killing, wherein full or partial suppression by a modifier allele is equivalent. **B)** Post-meiotic killing, under the scenarios of a fully or partially suppressing modifier. **C)** Meso-meiotic killing, under the scenarios of a fully or partially suppressing modifier.

A)

Haplotype	Pre-meiotic; Full/Partial suppression
DM	1
Dm	$1 - f_m s + f_d f_m e(1-s)$
dM	1
dm	$1 - f_m s - f_D f_m e(1-s)$

B)

Haplotype	Post-meiotic; Full suppression	Post-meiotic; Partial suppression
DM	$1 - f_m\big(s + e(1-s)\big)$	$1 - (1 - f_D f_M)s + f_d f_m e(1-s)$
Dm	$1 - s + (f_d + f_M + f_d f_M)e(1-s)$	$1 - s + f_d e(1-s)$
dM	$1 - f_D f_m\big(s + e(1-s)\big)$	$1 - f_D f_m\big(s + e(1-s)\big)$
dm	$1 - f_D\big(s + e(1-s)\big)$	$1 - f_D\big(s + e(1-s)\big)$

B)

Haplotype	Meso-meiotic; Full suppression	Meso-meiotic; Partial suppression
DM	$1 - f_D f_m\big(s + e(1-s)\big)$	$1 - f_D f_m s$
Dm	$1 - s + (1 - f_D f_m)e(1-s)$	$1 - s + f_d e(1-s)$
dM	$1 - f_D f_m\big(s + e(1-s)\big)$	$1 - f_D f_m\big(s + e(1-s)\big)$
dm	$1 - f_D f_m\big(s + e(1-s)\big)$	$1 - f_D f_m\big(s + e(1-s)\big)$

To describe how natural selection affects linkage disequilibrium, we first calculate the fitness of the four haplotypes ($\omega_{DM}, \omega_{Dm}, \omega_{dM}, \omega_{dm}$; see Table S1.1). The fitness expressions are complicated and so, to simplify the analysis of linkage disequilibrium to an informative level, we restrict our attention to describe the quantity of linkage disequilibrium that natural selection builds in a single generation from the starting condition where there is linkage equilibrium (often called D_0). This restriction simplifies the analysis without any loss of generality (because it allows us to identify and understand the phenomena that favour linkage disequilibrium while simply avoiding the complicating impacts of secondary, higher-order processes that arise once there is existing linkage disequilibrium). The linkage disequilibrium built by selection can be calculated using the haplotype frequencies after selection (see eqn. S1.2; *i.e.* $D = (f_D f_M \omega_{DM}/\bar{\omega})(f_d f_m \omega_{dm}/\bar{\omega}) - (f_D f_m \omega_{Dm}/\bar{\omega})(f_d f_M \omega_{dM}/\bar{\omega})$). To evaluate the evolution of linkage disequilibrium in the five scenarios of gamete-killing gene drive we use the definition of fitness given in Table S1.1. The expressions of linkage disequilibrium (Table S1.2) are complicated functions, hence we ignore interference and focus on the epistatic effect.

Across scenarios, the properties of the epistatic effect remain broadly consistent: under no scenario is epistasis absent ($E \neq 0$), the sign of the epistatic effect is always negative when the drive allele is spreading to fixation, and the size of the epistatic effect is frequency dependent. For pre-meiotic killing (under full or partial suppression, which are equivalent) and post-meiotic killing (under full or partial suppression, which are not equivalent), the epistatic effect is always negative ($E < 0$), which implies that selection favours the 'repulsion' allelic combinations (*i.e.* haplotypes: Dm, dM). The strength of epistatic selection depends on allele frequencies, but this effect is opposite in these two scenarios: the size of the epistatic effect in pre-meiotic killing increases as a function of the frequency of the nonmodifier allele (f_m), whilst for post-meiotic killing (especially under partial suppression) the size of the epistatic effect decreases as a function of the frequency of the nonmodifier allele (f_m; and hence increases as a function of the modifier allele, f_M). Post-meiotic killing under full

suppression is also influenced by the frequency of the drive allele (f_D), with the size of epistatic effect increasing with the frequency of the nondrive allele (f_d). Like other scenarios, though here because of the constraint $(1+e)(1-s)>1$ for the drive allele to be favoured (see main-text), the epistatic effect under meso-meiotic killing always has a negative sign, and the size of the epistatic effect is also frequency-dependent: the size of the epistatic effect in meso-meiotic killing increases as a function of the frequency of the unmodified drive haplotype ($f_D f_m$). Therefore, across different gamete-killing scenarios, natural selection tends to build linkage disequilibrium in the same way: the negative epistatic effect builds linkage disequilibrium because selection favours the 'repulsion' allelic combinations (*i.e.* haplotypes: Dm, dM). This conclusion is unsurprising because, across scenarios, the dm haplotype is vulnerable to the drive allele's toxin, which benefits the toxin-producing Dm haplotype.

Table S1.2. Linkage disequilibrium under different scenarios of gamete-killing, and its relationship to the epistatic effect. Linkage disequilibrium (D) built by natural selection after a single generation, starting from linkage equilibrium is calculated as $D = (f_D f_M \omega_{DM}/\bar{\omega})(f_d f_m \omega_{dm}/\bar{\omega}) - (f_D f_m \omega_{Dm}/\bar{\omega})(f_d f_M \omega_{dM}/\bar{\omega})$ (see eqn. S1.2). The epistatic effect is shown in red within the full expression for linkage disequilibrium: $E = (\omega_{DM} + \omega_{dm}) - (\omega_{Dm} + \omega_{Dm})$ (see eqn. S1.4).

Genotype	Linkage disequilibrium (D)
Pre-meiotic; Full/Partial suppression	$f_D f_d f_M f_m(-f_m e(1-s))/\bar{\omega}^2$
Post-meiotic; Full suppression	$f_D f_d f_M f_m \Big(f_d f_M s - (1+f_d+f_M)e(1-s) + f_D f_m(1+f_d)(1+f_M)e(1-s)\big(s+e(1-s)\big)\Big)/\bar{\omega}^2$
Post-meiotic; Partial suppression	$f_D f_d f_M f_m \Big(-f_M e(1-s) + f_D f_d f_M s\big(s+e(1-s)\big)\Big)/\bar{\omega}^2$
Meso-meiotic; Full suppression	$f_D f_d f_M f_m\big((1-f_D f_m)s - e(1-s)\big)\Big(1 - f_D f_m\big(s+e(1-s)\big)\Big)/\bar{\omega}^2$
Meso-meiotic; Partial suppression	$f_D f_d f_M f_m\big((1-f_D f_m)s - f_d e(1-s)\big)\Big(1 - f_D f_m\big(s+e(1-s)\big)\Big)/\bar{\omega}^2$

How large is the size of linkage disequilibrium? Across multilocus systems, linkage disequilibrium arrives at a dynamic equilibrium between new linkage disequilibrium that is built by natural selection and existing linkage disequilibrium that is eroded by recombination. Numerous studies have shown that recombination is efficient at removing linkage disequilibrium unless genes are genetically linked by close physical proximity along the chromosome (*e.g.* Geiringer, 1944; Lewontin, 1964; see also Lewontin, 1974). As our model supposes the general case of a modifier elsewhere in the genome, we have no reason to assume physical proximity so most modifiers are likely to be unlinked, implying that recombination rate is close to its maximum rate for the erosion of linkage disequilibrium (of 50% per generation, or 0.5). Consequently, the level of linkage disequilibrium that can be maintained at a dynamic equilibrium is likely to be small. However, because gene drives can lead to a rapid selective sweep, the nonequilibrium level of linkage disequilibrium could be larger as a drive or modifier allele sweeps to fixation leading linkage disequilibrium to be temporarily built up faster than recombination erodes. This can be seen within the expressions of linkage disequilibrium, which could be relatively large (compared to 0.5) when there is a larger transmission effect (e) and a smaller toxicity cost (s). Given that a drive allele is favoured when $(1 + e)(1 - s) > 1$ in the absence of a modifier allele, the choice of viable values for the transmission effect and toxicity cost in gene drive design are constrained, but (from the main-text) an effective gene drive would have a large transmission effect and a large toxicity cost because the magnitude of population suppression depends on the size of the toxicity cost (which is then balanced to make $(1 + e)(1 - s) > 1$ by having a larger transmission effect). This choice amounts to picking model parameters that lead to weaker selection (see Nagylaki 1993), increasing the time that it takes for the drive allele to spread to fixation in a nonmodifier allele background and reducing the build-up of linkage disequilibrium. Therefore, overall, the size of linkage disequilibrium that is built by natural selection in our model is dependent upon the transmission effect and toxicity cost parameters that determine the rapidity that either a drive allele spreads to fixation; given the main-text's argument for a larger transmission effect

and toxicity cost, the amount of linkage disequilibrium is likely to be relatively small compared to plausible rates of recombination.

Does linkage disequilibrium suggest alternative outcomes for natural selection? We should keep in mind that our analysis in the main-text suggests that either a modifier allele for full or partial suppression would successfully silence toxin production under pre-meiotic or post-meiotic killing, which prevents natural selection from building linkage disequilibrium because all alleles have equal fitness once a modifier is at fixation. So the purpose of introducing these scenarios was not to suggest that linkage disequilibrium might not be important, and in fact we have shown that natural selection favours linkage disequilibrium, but in these scenarios linkage disequilibrium only hastens the silencing of toxin production because it makes the nonmodifier allele have an even lower fitness (due to, by the negative sign of the epistatic effect, being more likely to be paired with the nondrive allele) – generally favouring the even more rapid invasion of a modifier allele (that is more likely to be paired with a drive allele). In this way, the additional complexity of incorporating linkage disequilibrium does not change the outcomes for pre-meiotic and post-meiotic killing. The meso-meiotic killing scenarios are more complicated. The size of the epistatic effect is smaller in these scenarios, especially when the drive and nonmodifier alleles are rare (see Table S1.2). This follows logically because the fitness of all haplotypes except the unmodified drive haplotype (Dm) are near (if not exactly) equal, so linkage disequilibrium is only arising because of the fitness advantage to the unmodified drive haplotype. Consequently, the linkage disequilibrium favours the more rapid invasion of the drive allele. When meso-meiotic killing is evaluated with a partially suppressing modifier, the more rapid invasion of the drive allele ultimately creates the conditions that lead to the more rapid invasion of a partially suppressing modifier (that arises by mutation in a drive allele background), which reverses the sign of the epistatic effect such that selection favours the 'coupling' allelic combinations (*i.e.* haplotypes: DM, dm) because of the fitness advantage of the modified drive haplotype (DM). Therefore, despite arising through slightly different means in different scenarios, linkage disequilibrium

does not suggest alternative outcomes for natural selection, but it does increase the speed of reaching a stable equilibrium for the drive and modifier loci.

In summary, our simple analysis here has shown that linkage disequilibrium can play a role in gamete-killing gene drive. Under the desirable model parameters for the design of effective gene drives suggested in the main-text, the size of linkage disequilibrium is likely to be relatively small compared to plausible recombination rates, preventing a dynamically stable equilibrium with linkage disequilibrium. Further, linkage disequilibrium that does arise only increases the speed at which evolutionary changes in gene frequency occur under natural selection, rather than having the potential to change the outcome of natural selection. As a result, there is no need to consider a more complicated model of the five scenarios under linkage disequilibrium to justify the conclusions in the main-text. However, the simple analysis of linkage disequilibrium presented here suggests that, were gamete-killing to be taken forward as a mechanism of gene drive, linkage disequilibrium would likely play some role – and so should be considered.

References

Geiringer, H. 1944. On the Probability Theory of Linkage in Mendelian Heredity. *Ann. Math. Stat.* **15**: 25–57.

Hallgrímsdóttir, I.B. & Yuster, D.S. 2008. A complete classification of epistatic two-locus models. *BMC Genet.* **9**: 1–15.

Hartl, D.L. & Clark, A.G. 2006. *Principles of Population Genetics*. Sinauer Associates.

Hastings, A. 1985. Multilocus population genetics with weak epistasis. I. Equilibrium properties of two-locus two-allele models. *Genetics* **109**: 799–812.

Lewontin, R.C. 1974. *The genetic basis of evolutionary change*. Columbia University Press.

Lewontin, R.C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**: 49–67.

Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* **134**: 627–647.

Bibliography

Abbot, P., Withgott, J.H. & Moran, N.A. 2001. Genetic conflict and conditional altruism in social aphid colonies. *Proc. Natl. Acad. Sci.* **98**: 12068–12071.

Ågren, J.A., Davies, N.G. & Foster, K.R. 2019. Enforcement is central to the evolution of cooperation. *Nat. Ecol. Evol.* **3**: 1018–1029. Springer US.

Alexander, R. & Borgia, G. 1978. Group Selection, Altruism, and the Levels of Organization of Life. *Annu. Rev. Ecol. Syst.* **9**: 449–474.

Axelrod, R. 1984. *The Evolution of Cooperation*. Basic Books, New York.

Axelrod, R. & Hamilton, W.D. 1981. The evolution of cooperation. *Science (80-.).* **211**: 1390–1396.

Axelrod, R., Hammond, R.A. & Grafen, A. 2004. Altruism via kin-selection strategies that rely on arbitrary tags with which they coevolve. *Evolution* **58**: 1833–1838.

Bachrach, P. & Baratz, M.S. 1962. Two Faces of Power. *Am. Polit. Sci. Rev.* **56**: 1962.

Beeman, R.W., Friesen, K.S. & Denell, R.E. 1992. Maternal-effect selfish genes in flour beetles. *Science (80-.).* **256**: 89–92.

Ben-Zion, I., Pollak, S. & Eldar, A. 2018. Clonality and non-linearity drive facultative-cooperation allele diversity. *ISME J.* **13**: 824–835.

Benabentos, R., Hirose, S., Sucgang, R., Curk, T., Katoh, M., Ostrowski, E.A., *et al.* 2009. Polymorphic members of the lag gene family mediate kin discrimination in Dictyostelium. *Curr. Biol.* **19**: 567–572.

Bergstrom, T., Blume, L. & Varian, H. 1986. On the private provision of public goods. *J. Public Econ.* **29**: 25–49.

Biernaskie, J.M., Gardner, A. & West, S.A. 2013. Multicoloured greenbeards, bacteriocin diversity and the rock-paper-scissors game. *J. Evol. Biol.* **26**: 2081–2094.

Biernaskie, J.M., West, S.A. & Gardner, A. 2011. Are Greenbeards Intragenomic Outlaws? *Evolution (N. Y).* **65**: 2729–2742.

Binmore, K. 2005. *Natural Justice*. Oxford University Press.

Bourke, A.F.G. 2011. *Principles of Social Evolution*. Oxford University Press.

Boyd, R., Gintis, H. & Bowles, S. 2010. Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science (80-.).* **328**: 617–620.

Boyd, R. & Richerson, P.J. 1988. The Evolution of Reciprocity in Sizable Groups. *J. Theor. Biol.* **132**: 337–356.

Bravo Núñez, M.A., Nuckolls, N.L. & Zanders, S.E. 2018. Genetic Villains: Killer Meiotic Drivers. *Trends Genet.* **34**: 424–433.

Breed, M.D., Welch, C.K. & Cruz, R. 1994. Kin discrimination within honey bee (Apis mellifera) colonies: An analysis of the evidence. *Behav. Processes* **33**: 25–39.

Bruce, J.B., Cooper, G.A., Chabas, H., West, S.A. & Griffin, A.S. 2017. Cheating and resistance to cheating in natural populations of the bacterium Pseudomonas fluorescens. *Evolution (N. Y).* **71**: 2484–2495.

Bshary, R. & Grutter, A.S. 2005. Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biol. Lett.* **1**: 396–399.

Bshary, R., Grutter, A.S., Willener, A.S.T. & Leimar, O. 2008. Pairs of cooperating cleaner fish provide better service quality than singletons. *Nature* **455**: 964–966.

Buchman, A., Marshall, J.M., Ostrovski, D., Yang, T. & Akbari, O.S. 2018. Synthetically engineered Medea gene drive system in the worldwide crop pest Drosophila suzukii. *Proc. Natl. Acad. Sci.* **115**: 4725–4730.

Bull, J.J. & Malik, H.S. 2017. The gene drive bubble: New realities. *PLoS Genet.* **13**: e1006850.

Bull, J.J. & Rice, W.R. 1991. Distinguishing mechanisms for the evolution of co-operation. *J. Theor. Biol.* **149**: 63–74.

Burt, A. 2014. Heritable strategies for controlling insect vectors of disease. *Philos. Trans. R. Soc. B Biol. Sci.* **369**: 20130432.

Burt, A. 2003. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. *Proc. R. Soc. B Biol. Sci.* **270**: 921–928.

Burt, A. & Trivers, R. 2008. *Genes in Conflict*. Harvard University Press.

Buss, L.W. 1982. Somatic cell parasitism and the evolution of somatic tissue compatibility. *Proc. Natl. Acad. Sci.* **79**: 5337–5341.

Buttery, N.J., Rozen, D.E., Wolf, J.B. & Thompson, C.R.L. 2009. Quantification of social behavior in D. discoideum reveals complex fixed and facultative strategies. *Curr. Biol.* **19**: 1373–1377.

Cadavid, L.F., Powell, A.E., Nicotra, M.L., Moreno, M. & Buss, L.W. 2004. An invertebrate histocompatibility complex. *Genetics* **167**: 357–365.

Cao, P. & Wall, D. 2017. Self-identity reprogrammed by a single residue switch in a cell surface receptor of a social bacterium. *Proc. Natl. Acad. Sci.* **114**: 3732–3737.

Cao, P., Wei, X., Awal, R.P., Müller, R. & Wall, D. 2019. A Highly Polymorphic Receptor Governs Many Distinct Self-Recognition Types within the Myxococcales Order. *MBio* **10**: 1–15.

Cazemajor, M., Joly, D. & Montchamp-Moreau, C. 2000. Sex-ratio meiotic drive in Drosophila simulans is related to equational nondisjunction of the Y chromosome. *Genetics* **154**: 229–236.

Champer, J., Buchman, A. & Akbari, O.S. 2016. Cheating evolution: Engineering gene drives to manipulate the fate of wild populations. *Nat. Rev. Genet.* **17**: 146–159.

Champer, J., Liu, J., Oh, S.Y., Reeves, R., Luthra, A., Oakes, N., *et al.* 2018. Reducing resistance allele formation in CRISPR gene drive. *Proc. Natl. Acad. Sci.* **115**: 5522–5527.

Champer, J., Reeves, R., Oh, S.Y., Liu, C., Liu, J., Clark, A.G., *et al.* 2017. Novel CRISPR/Cas9 gene drive constructs reveal insights into mechanisms of resistance allele formation and drive efficiency in genetically diverse populations. *PLoS Genet.* **13**: 1–18.

Charlesworth, B. & Hartl, D.I. 1978. Population dynamics of the segregation distorter polymorphism of Drosophila melanogaster. *Genetics* **89**: 171–192.

Charnov, E.L. 1977. An elementary treatment of the genetical theory of kin-selection. *J. Theor. Biol.* **66**: 541–550.

Chattwood, A., Nagayama, K., Bolourani, P., Harkin, L., Kamjoo, M., Weeks, G., *et al.* 2013. Developmental lineage priming in Dictyostelium by heterogeneous Ras activation. *Elife* **2**: e01067.

Chen, C.H., Huang, H., Ward, C.M., Su, J.T., Schaeffer, L. V., Guo, M., *et al.* 2007. A synthetic maternal-effect selfish genetic element drives population replacement in Drosophila. *Science (80-.).* **316**: 597–600.

Clutton-Brock, T. 2009. Cooperation between non-kin in animal societies. *Nature* **462**: 51–57.

Clutton-Brock, T.H. & Parker, G.A. 1995. Punishment in animal societies. *Nature* **373**: 209–216.

Clutton-Brock, T.H., Riain, M.J.O. & Brotherton, P.N.M. 1999. Selfish Sentinels in Cooperative Mammals. *Science (80-.).* **284**: 1640–1644.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. *J. Theor. Biol.* **12**: 119–129.

Cornwallis, C.K., West, S.A. & Griffin, A.S. 2009. Routes to indirect fitness in cooperatively breeding vertebrates: Kin discrimination and limited dispersal. *J. Evol. Biol.* **22**: 2445–2457.

Cosmides, L.M. & Tooby, J. 1981. Cytoplasmic inheritance and intragenomic conflict. *J. Theor. Biol.* **89**: 83–129.

Crampton, W.G. & Hurst, L.D. 1994. True kin recognition, in the form of somatic incompatibility, has multiple independent origins. *Anim. Behav.* **47**: 230–234.

Crozier, R. 1986. Genetic clonal recognition abilities in marine invertebrates must be maintained by selection for something else. *Evolution (N. Y).* **40**: 1100–1101.

Danka, E.S., Garcia, E.C. & Cotter, P.A. 2017. Are CDI Systems Multicolored, Facultative, Helping Greenbeards? *Trends Microbiol.* **25**: 391–401.

Dawkins, R. 1982. *The Extended Phenotype: The Gene as the Unit of Selection*. Oxford University Press.

Dawkins, R. 1976. *The Selfish Gene*. Oxford University Press.

Dawkins, R. 1979. Twelve misunderstandings of kin selection. *Z. Tierpsychol.* **51**: 184–200.

de Oliveira, J.L., Morales, A.C., Stewart, B., Gruenheit, N., Engelmoer, J., Brown, S.B., *et al.* 2019. Conditional expression explains molecular evolution of social genes in a microbe. *Nat. Commun.* **10**: 3284. Springer US.

De Tomaso, A.W. 2018. Allorecognition and Stem Cell Parasitism: A Tale of Competition, Selfish Genes and Greenbeards in a Basal Chordate. In: *Origin and Evolution of Biodiversity* (P. Pontarotti, ed), pp. 131–142. Springer International Publishing.

Deredec, A., Burt, A. & Godfray, H.C.J. 2008. The population genetics of using homing endonuclease genes in vector and pest management. *Genetics* **179**: 2013–2026.

Dickinson, J.L. 2004. A test of the importance of direct and indirect fitness benefits for helping decisions in western bluebirds. *Behav. Ecol.* **15**: 233–238.

Diggle, S.P., Griffin, A.S., Campbell, G.S. & West, S.A. 2007. Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**: 411–414.

Dionisio, F. & Gordo, I. 2007. Controlling excludability in the evolution of cooperation. *Evol. Ecol. Res.* **9**: 365–373.

Dionisio, F. & Gordo, I. 2006. The tragedy of the commons, the public good dilemma and the meaning of rivalry and excludability in evolutionary biology. *Evol. Ecol. Res.* **8**: 321–332.

Doebeli, M. & Hauert, C. 2005. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* **8**: 748–766.

Doebeli, M., Hauert, C. & Killingback, T. 2004. The Evolutionary Origin of Cooperators and Defectors. *Science (80-.).* **306**: 859–862.

Drescher, K., Nadell, C.D., Stone, H.A., Wingreen, N.S. & Bassler, B.L. 2014. Solutions to the public goods dilemma in bacterial biofilms. *Curr. Biol.* **24**: 50–55.

Dugatkin, L.A. & Reeve, H.K. 1998. *Game Theory and Animal Behavior*. Oxford University Press.

Duncan, C., Gaynor, D., Clutton-Brock, T. & Dyble, M. 2019. The Evolution of Indiscriminate Altruism in a Cooperatively Breeding Mammal. *Am. Nat.* **193**: 841–851.

Eshel, I. 1985. Evolutionary Genetic Stability of Mendelian Segregation and the Role of Free Recombination in the Chromosomal System. *Am. Nat.* **125**: 412–420.

Faria, G.S., Varela, S.A.M. & Gardner, A. 2018. The relation between R . A . Fisher's sexy-son hypothesis and W . D . Hamilton's greenbeard. *Evol. Lett.* **2**: 190–200.

Ferrari, M. & König, B. 2017. No evidence for punishment in communally nursing female house mice (Mus musculus domesticus). *PLoS One* **12**: 1–16.

Ferrari, M., Lindholm, A.K. & König, B. 2015. The risk of exploitation during communal nursing in house mice, Mus musculus domesticus. *Anim. Behav.* **110**: 133–143.

Ferrari, M., Lindholm, A.K., König, B. & Ko, B. 2016. A reduced propensity to cooperate under enhanced exploitation risk in a social mammal. *Procedings R. Soc. B Biol. Sci.* **283**: 20160068.

Flood, M.M. 1952. Some Experimental Games. *RAND Corp.* **RM-789-1**: 1–44.

Forman, D. & Garrod, D.R. 1977. Pattern formation in Dictyostelium discoideum. I. Development of prespore cells and its relationship to the pattern of the fruiting body. *J. Embryol. Exp. Morphol.* **Vol. 40**: 215–228.

Foster, K., Wenseleers, T. & Ratnieks, F. 2006. Kin selection is the key to altruism. *Trends Ecol. Evol.* **21**: 57–60.

Foster, K.R. 2009. A defense of sociobiology. *Cold Spring Harb. Symp. Quant. Biol.* **74**: 403–18.

Foster, K.R. 2004. Diminishing returns in social evolution: The not-so-tragic commons. *J. Evol. Biol.* **17**: 1058–1072.

Foster, K.R., Fortunato, A., Strassmann, J.E. & Queller, D.C. 2002. The costs and benefits of being a chimera. *Proc. Biol. Sci.* **269**: 2357–2362.

Foster, K.R., Shaulsky, G., Strassmann, J.E., Queller, D.C. & Thompson, C.R.L. 2004. Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**: 693–696.

Fox, J. & Guyer, M. 1978. "Public" Choice and Cooperation in n-Person Prisoner's Dilemma. *J. Conflict Resolut.* **22**: 469–481.

Frank, S.A. 2010. A general model of the public goods dilemma. *J. Evol. Biol.* **23**: 1245–1250.

Frank, S.A. 2009. Evolutionary Foundations of Cooperation and Group Cohesion. In: *Games,*

Groups, and the Global Good (S. Levin, ed), pp. 3–40. Springer Physica-Verlag Berlin Heidelberg 2009.

Frank, S.A. 1998. *Foundations of Social Evolution*. Princeton University Press.

Frank, S.A. 1994. Genetics of mutualism: the evolution of altruism between species. *J. Theor. Biol.* **170**: 393–400.

Frank, S.A. 1995. Mutual policing and repression of competition in the evolution of cooperative groups. *Nature* **377**: 520–522.

Frank, S.A. 2013. Natural selection. VII. History and interpretation of kin selection theory. *J. Evol. Biol.* **26**: 1151–1184.

Frank, S.A. 1996. Policing and group cohesion when resources vary. *Anim. Behav.* **52**: 1163–1169.

Frank, S.A. 2003. Repression of competition and the evolution of cooperation. *Evolution (N. Y).* **57**: 693–705.

Frank, S.A. 1987. Variable sex ratio among colonies of ants. *Behav. Ecol. Sociobiol.* **20**: 195–201.

Gardner, A. & Grafen, A. 2009. Capturing the superorganism: A formal theory of group adaptation. *J. Evol. Biol.* **22**: 659–671.

Gardner, A. & Úbeda, F. 2017. The meaning of intragenomic conflict. *Nat. Ecol. Evol.* **1**: 1807–1815.

Gardner, A. & West, S. 2014. Inclusive fitness: 50 years on. *Phil. Trans. R. Soc. B* **369**: 1–3.

Gardner, A. & West, S.A. 2004a. Cooperation and Punishment, Especially in Humans. *Am. Nat.* **164**: 753–764.

Gardner, A. & West, S.A. 2010. Greenbeards. *Evolution (N. Y).* **64**: 25–38.

Gardner, A. & West, S.A. 2004b. Spite and the scale of competition. *J. Evol. Biol.* **17**: 1195–1203.

Gavrilets, S. 2015. Collective action problem in heterogeneous groups. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **370**: 20150016.

Gaynor, K.M. & Cords, M. 2012. Antipredator and social monitoring functions of vigilance behaviour in blue monkeys. *Anim. Behav.* **84**: 531–537.

Geiringer, H. 1944. On the Probability Theory of Linkage in Mendelian Heredity. *Ann. Math. Stat.* **15**: 25–57.

Ghoul, M., Griffin, A.S. & West, S.A. 2014. Toward an evolutionary definition of cheating. *Evolution (N. Y).* 318–331.

Gilbert, O.M., Foster, K.R., Mehdiabadi, N.J., Strassmann, J.E. & Queller, D.C. 2007. High relatedness maintains multicellular cooperation in a social amoeba by controlling cheater mutants. *Proc. Natl. Acad. Sci.* **104**: 8913–7.

Gilbert, O.M., Strassmann, J.E. & Queller, D.C. 2012. High relatedness in a social amoeba: the role of kin-discriminatory segregation. *Proc. R. Soc. B Biol. Sci.* **279**: 2619–2624.

Godfray, H.C.J., North, A. & Burt, A. 2017. How driving endonuclease genes can be used to combat pests and disease vectors. *BMC Biol.* **15**: 1–12.

Gordon, H.S. 1954. The Economic Theory of a Common-Property Resource: The Fishery. *J.*

Polit. Econ. **62**: 124–142.

Gore, J., Youk, H. & van Oudenaarden, A. 2009. Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**: 253–6.

Grafen, A. 1985. A geometric view of relatedness. *Oxford Surv. Evol. Biol.* **2**: 28–89.

Grafen, A. 1990. Do animals really recognize kin? *Anim. Behav.* **39**: 42–54.

Grafen, A. 1984. Natural selection, kin selection and group selection. In: *Behavioural Ecology* (J. R. Krebs & N. B. Davies, eds), pp. 62–84. Blackwell Science Ltd.

Grafen, A. 2006a. Optimization of inclusive fitness. *J. Theor. Biol.* **238**: 541–563.

Grafen, A. 2006b. Various remarks on Lehmann and Keller's article. *J. Evol. Biol.* **19**: 1397–1399.

Griffin, A.S., West, S.A. & Buckling, A. 2004. Cooperation and competition in pathogenic bacteria. *Nature* **430**: 1024–1027.

Gruenheit, N., Parkinson, K., Stewart, B., Howie, J., Wolf, J. & Thomposon, C. 2017. A polychromatic "greenbeard" locus determines patterns of cooperation in a social amoeba. *Nat. Commun.* **14171**: 1–9.

Guilford, T. 1985. Is kin selection involved in the evolution of warning coloration? *Oikos* **45**: 31–36.

Guilford, T. 1988. The Evolution of Conspicuous Coloration. *Am. Nat.* **131**: S7–S21.

Haig, D. 1996. Gestational drive and the green-bearded placenta. *Proc. Natl. Acad. Sci.* **93**: 6547–6551.

Haig, D. 2013. Imprinted green beards: a little less than kin and more than kind. *Biol. Lett.* **9**: 20130199.

Haig, D. 1997. The social gene. In: *Behavioural Ecology: an evolutionary approach* (J. R. Krebs & N. B. Davies, eds), pp. 284–304. Blackwell Science Ltd.

Haig, D. & Grafen, A. 1991. Genetic scrambling as a defence against meiotic drive. *J. Theor. Biol.* **153**: 531–558.

Haldane, J.B.S. 1955. Population Genetics. *New Biol.* **18**: 34.

Haldane, J.B.S. 1932. *The Causes of Evolution*. Longmans, Green and Co. Limited.

Hallgrímsdóttir, I.B. & Yuster, D.S. 2008. A complete classification of epistatic two-locus models. *BMC Genet.* **9**: 1–15.

Hamburger, H. 1973. N-person Prisoner's Dilemma. *J. Math. Sociol.* **3**: 27–48.

Hamilton, W.D. 1972. Altruism and Related Phenomena, Mainly in Social Insects. *Annu. Rev. Ecol. Syst.* **3**: 193–232.

Hamilton, W.D. 1975. Innate Social Aptitudes of Man: an approach from evolutionary genetics. In: *ASA Studies 4: Biosocial Anthropology* (R. Fox, ed), pp. 133–153. Malaby Press.

Hamilton, W.D. 2001. *Narrow Roads of Gene Land: Volume 2*. Oxford University Press.

Hamilton, W.D. 1971. Selection of selfish and altruistic behavior in some extreme models. In: *Man and beast: comparative social behavior* (J. F. Eisenberg & W. S. Dillon, eds), pp. 57–91. Smithsonian Institution Press.

Hamilton, W.D. 1963. The Evolution of Altruistic Behavior. *Am. Nat.* **97**: 354–356.

Hamilton, W.D. 1964a. The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**: 17–52.

Hamilton, W.D. 1964b. The genetical evolution of social behaviour II. *J. Theor. Biol.* **7**: 17–52.

Hammond, A.M., Kyrou, K., Bruttini, M., North, A., Galizi, R., Karlsson, X., *et al.* 2017. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. *PLoS Genet.* **13**: e1007039.

Hardin, G. 1968. The tragedy of the commons. *Science (80-.).* **162**: 1243–1248.

Hartl, D.L. 1975. Modifier theory and meiotic drive. *Theor. Popul. Biol.* **7**: 168–174.

Hartl, D.L. & Clark, A.G. 2006. Principles of Population Genetics. Sinauer Associates.

Hastings, A. 1985. Multilocus population genetics with weak epistasis. I. Equilibrium properties of two-locus two-allele models. *Genetics* **109**: 799–812.

Heller, J., Zhao, J., Rosenfield, G., Kowbel, D.J., Gladieux, P. & Glass, N.L. 2016. Characterization of Greenbeard Genes Involved in Long-Distance Kind Discrimination in a Microbial Eukaryote. *PLoS Biol.* **14**: e1002431.

Helleu, Q., Gérard, P.R., Dubruille, R., Ogereau, D., Prud'homme, B., Loppin, B., *et al.* 2016. Rapid evolution of a Y-chromosome heterochromatin protein underlies sex chromosome meiotic drive. *Proc. Natl. Acad. Sci.* **113**: 4110–4115.

Hirose, S., Benabentos, R., Ho, H.-I., Kuspa, A. & Shaulsky, G. 2011. Self-Recognition in Social Amoebae Is Mediated by Allelic Pairs of Tiger Genes. *Science (80-.).* **333**: 467–470.

Ho, H.-I. & Shaulsky, G. 2015. Temporal regulation of kin recognition maintains recognition-cue diversity and suppresses cheating. *Nat. Commun.* **6**: 7144.

Hornett, E.A., Charlat, S., Wedell, N., Jiggins, C.D. & Hurst, G.D.D. 2009. Rapidly Shifting Sex Ratio across a Species Range. *Curr. Biol.* **19**: 1628–1631.

Huang, Y.C. & Wang, J. 2014. Did the fire ant supergene evolve selfishly or socially? *BioEssays* **36**: 200–208.

Hurst, G.D.D. & McVean, G.A.T. 1998. Selfish genes in a social insect. *Trends Ecol. Evol.* **13**: 434–435.

Hurst, L.D.., Atlan, A. & Bentsson, B.O. 1996. Genetic Conflicts. *Q. Rev. Biol.* **71**: 317–364.

Jansen, V.A.A. & van Baalen, M. 2006. Altruism through beard chromodynamics. *Nature* **440**: 663–666.

Johnstone, R.A. & Rodrigues, A.M.M. 2016. Cooperation and the common good. *Phil Trans R Soc B* **371**: 20150086.

Karadge, U.B., Gosto, M. & Nicotra, M.L. 2015. Allorecognition Proteins in an Invertebrate Exhibit Homophilic Interactions. *Curr. Biol.* **25**: 2845–2850.

Keller, L. 1997. Indiscriminate altruism: Unduly nice parents and siblings. *Trends Ecol. Evol.* **12**: 99–103.

Keller, L. 2002. Marsupial sperm pairing: A case of "sticky" green beards? *Trends Ecol. Evol.* **17**: 112–113.

Keller, L. & Ross, K.G. 1998. Selfish genes: a green beard in the red fire ant. *Nature* **394**: 573–575.

Kennedy, P., Higginson, A.D., Radford, A.N. & Sumner, S. 2018. Altruism in a volatile world. *Nature* **555**: 359–362. Nature Publishing Group.

Kessin, R.H. 2001. *Dictyostelium: evolution, cell biology, and the development of multicellularity*. Cambridge University Press.

Khare, A. & Shaulsky, G. 2010. Cheating by exploitation of developmental prestalk patterning in Dictyostelium discoideum. *PLoS Genet.* **6**: e1000854.

Kirkpatrick, R. & Gomulkiewicz, M. 1992. Quantitative Genetics and the Evolution of Reaction Norms. *Evolution (N. Y).* **46**: 390–411.

Komdeur, J., Richardson, D.S. & Burke, T. 2004. Experimental evidence that kin discrimination in the Seychelles warbler is based on association and not on genetic relatedness. *Proc. R. Soc. B Biol. Sci.* **271**: 963–969.

Konig, B. 1994. Fitness effects of communal rearing in house mice: the role of relatedness versus familiarity. *Anim. Behav.* **48**: 1449–1457.

Kovacevic, M. & Schaeffer, S.W. 2000. Molecular population genetics of X-linked genes in Drosophila pseudoobscura. *Genetics* **156**: 155–172.

Krupp, D.B. & Taylor, P.D. 2015. Social evolution in the shadow of asymmetrical relatedness. *Proc. R. Soc. B Biol. Sci.* **282**: 20150142.

Leigh, E.G. 1971. *Adaptation and Diversity: natural history and the mathematics of evolution*. Freeman, Cooper.

Leigh, E.G. 1977. How does selection reconcile individual advantage with the good of the group? *Proc. Natl. Acad. Sci.* **74**: 4542–4546.

Lewis, D. 1969. *Convention*. Harvard University Press.

Lewontin, R.C. 1974. *The genetic basis of evolutionary change*. Columbia University Press.

Lewontin, R.C. 1964. The Interaction of Selection and Linkage. I. General Considerations; Heterotic Models. *Genetics* **49**: 49–67.

Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., *et al.* 2016. The Ecology and Evolutionary Dynamics of Meiotic Drive. *Trends Ecol. Evol.* **31**: 315–326.

Linksvayer, T.A., Busch, J.W. & Smith, C.R. 2013. Social supergenes of superorganisms: do supergenes play important roles in social evolution? *Bioessays* **35**: 683–689.

Lyttle, T.W. 1977. Experimental population genetics of meiotic drive systems. I. Pseudo Y chromosomal drive as a means of eliminating cage populations of Drosophila melanogaster. *Genetics* **86**: 413–445.

Madgwick, P.G., Belcher, L.J., Stewart, B., Thompson, C.R.L. & Wolf, J.B. 2018. Strategic investment explains patterns of cooperation and cheating in a microbe. *Proc. Natl. Acad. Sci.* **115**: E4823–E4832.

Madgwick, P.G., Belcher, L.J. & Wolf, J.B. 2019. Greenbeard genes: theory and reality. *Trends Ecol. Evol.*

Manhes, P. & Velicer, G.J. 2011. Experimental evolution of selfish policing in social bacteria. *Proc. Natl. Acad. Sci.* **108**: 8357–8362.

Marshall, J.A.R. 2015. *Social Evolution and Inclusive Fitness Theory*. Princeton University Press.

Marshall, J.M. & Akbari, O.S. 2018. Can CRISPR-Based Gene Drive Be Confined in the Wild? A Question for Molecular and Population Biology. *ACS Chem. Biol.* **13**: 424–430.

Maynard Smith, J. 1976. Evolution and the Theory of Games. *Am. Sci.* **64**: 41–45. Cambridge University Press.

Maynard Smith, J. 1979. Game theory and the evolution of behaviour. *Proc. R. Soc. London. Ser. B, Biol. Sci.* **205**: 475–488. Edinburgh University Press.

Maynard Smith, J. 1964. Group Selection and Kin Selection. *Nature* **201**: 1145–1147.

Maynard Smith, J. 1982. Models of social behaviour - a classification of models. In: *Current Problems in Sociobiology* (King's College Sociobiology Group, ed). Cambridge University Press.

Maynard Smith, J. 1974. The theory of games and the evolution of animal conflicts. *J. Theor. Biol.* **47**: 209–221.

Maynard Smith, J. & Price, G.R. 1973. Logic of animal conflict. *Nature* **246**: 15–18.

McFarlane, G.R., Whitelaw, C.B.A. & Lillico, S.G. 2018. CRISPR-Based Gene Drives for Pest Control. *Trends Biotechnol.* **36**: 130–133.

McLaughlin, R.N. & Malik, H.S. 2017. Genetic conflicts: the usual suspects and beyond. *J. Exp. Biol.* **220**: 6–17.

McNamara, J.M. 2013. Towards a richer evolutionary game theory. *J. R. Soc. Interface* **10**.

Morris, H.R., Taylor, G.W., Masento, M.S., Jermyn, K.A. & Kay, R.R. 1987. Chemical structure of the morphogen differentiation inducing factor from Dictyostelium discoideum. *Nature* **328**: 811–814.

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

Nagylaki, T. 1993. The evolution of multilocus systems under weak selection. *Genetics* **134**: 627–647.

Nair, R.R., Fiegna, F. & Velicer, G.J. 2018. Indirect evolution of social fitness inequalities and facultative social exploitation. *Proc. R. Soc. B Biol. Sci.* **285**.

Nicotra, M.L., Powell, A.E., Rosengarten, R.D., Moreno, M., Grimwood, J., Lakkis, F.G., *et al.* 2009. A Hypervariable Invertebrate Allodeterminant. *Curr. Biol.* **19**: 583–589.

Noble, C., Adlam, B., Church, G.M., Esvelt, K.M. & Nowak, M.A. 2018. Current CRISPR gene drive systems are likely to be highly invasive in wild populations. *Elife* **7**: e33423.

Noble, C., Olejarz, J., Esvelt, K.M., Church, G.M. & Nowak, M.A. 2017. Evolutionary dynamics of CRISPR gene drives. *Sci. Adv.* **3**: 3–10.

Nonacs, P. 2011. Kinship, greenbeards, and runaway social selection in the evolution of social insect cooperation. *Proc. Natl. Acad. Sci.* **108**: 10808–10815. The National Academies Press.

Nowak, M.A. 2006. Fiver rules for the evolution of cooperation. *Science (80-.).* **314**: 1560–1563.

Nydam, M.L., Netuschil, N., Sanders, E., Langenbacher, A., Lewis, D.D., Taketa, D.A., *et al.* 2013. The Candidate Histocompatibility Locus of a Basal Chordate Encodes Two Highly

Polymorphic Proteins. *PLoS One* **8**: e65980.

Olofsson, H., Ripa, J. & Jonzén, N. 2009. Bet-hedging as an evolutionary game: The trade-off between egg size and number. *Proc. R. Soc. B Biol. Sci.* **276**: 2963–2969.

Olson, M. 1965. *The Logic of Collective Action*. Harvard University Press.

Ostrom, E. 1990. *Governing the Commons: the evolution of institutions for collective action*. Cambridge University Press, Cambridge.

Ostrowski, E.A., Katoh, M., Shaulsky, G., Queller, D.C. & Strassmann, J.E. 2008. Kin discrimination increases with genetic distance in a social amoeba. *PLoS Biol.* **6**: 2376–2382.

Palopoli, M.F. & Wu, C. 1996. Rapid Evolution of a Coadapted Gene Complex: Evidence From the Segregatuib Distorter (SD) System of Meiotic Drive in Drosophila melanogaster. *Genet. Soc. Am.* **143**: 1675–1688.

Parker, G.A. & Maynard Smith, J. 1990. Optimality theory in evolutionary biology. *Nature* **348**: 27–33.

Parkinson, K., Buttery, N.J., Wolf, J.B. & Thompson, C.R.L. 2011. A simple mechanism for complex social behavior. *PLoS Biol.* **9**: e1001039.

Pathak, D.T., Wei, X., Dey, A. & Wall, D. 2013. Molecular Recognition by a Polymorphic Cell Surface Receptor Governs Cooperative Behaviors in Bacteria. *PLoS Genet.* **9**: e1003891.

Pellmyr, O. & Huth, C.J. 1994. Evolutionary stability of mutualism between yuccas and yucca moth. *Nature* **372**: 257–261.

Pepper, J.W. 2000. Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**: 355–368.

Pepper, J.W. & Smuts, B.B. 2002. A mechanism for the evolution of altruism among nonkin: Positive assortment through environmental feedback. *Am. Nat.* **160**: 205–213.

Philippi, T. & Seger, J. 1989. Hedging One's Evolutionary Bets, Revisited. *Trends Ecol. Evol.* **4**: 2–5.

Pinzone, C.A. & Dyer, K.A. 2013. Association of polyandry and sex-ratio drive prevalence in natural populations of Drosophila neotestacea. *Proc. R. Soc. B Biol. Sci.* **280**: 20131397.

Pollak, S., Omer-Bendori, S., Even-Tov, E., Lipsman, V., Bareia, T., Ben-Zion, I., *et al.* 2016. Facultative cheating supports the coexistence of diverse quorum-sensing alleles. *Proc. Natl. Acad. Sci.* **113**: 2152–2157.

Poundstone, W. 1993. Prisoner's Dilemma. Anchor Books.

Pracana, R., Levantis, I., Martínez-Ruiz, C., Stolle, E., Priyam, A. & Wurm, Y. 2017. Fire ant social chromosomes: Differences in number, sequence and expression of odorant binding proteins. *Evol. Lett.* **1**: 199–210.

Price, T.A.R., Hurst, G.D.D. & Wedell, N. 2010. Polyandry Prevents Extinction. *Curr. Biol.* **20**: 471–475.

Prout, T., Bundgaard, J. & Bryant, S. 1973. Population genetics of modifiers of meiotic drive I. The solution of a special case and some general implications. *Theor. Popul. Biol.* **4**: 446–465.

Purcell, J., Brelsford, A., Wurm, Y., Perrin, N. & Chapuisat, M. 2014. Convergent genetic architecture underlies social organization in ants. *Curr. Biol.* **24**: 2728–2732.

Queller, D.C. 2011. Expanded social fitness and Hamilton's rule for kin, kith, and kind. *Proc. Natl. Acad. Sci.* **108**: 10792–10799.

Queller, D.C. 1994. Genetic relatedness in viscous populations. *Evol. Ecol.* **8**: 70–73.

Queller, D.C. 1984. Kin selection and frequency dependence: a game theoretic approach. *Biol. J. Linn. Soc.* **23**: 133–143.

Queller, D.C. 1992. Quantitative Genetics, Inclusive Fitness, and Group Selection. *Am. Nat.* **139**: 540–558.

Queller, D.C., Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F. & Smalley, R.E. 1985. Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**: 366–367.

Queller, D.C., Ponte, E., Bozzaro, S. & Strassmann, J.E. 2003. Single-Gene Greenbeard Effects in the Social Amoeba Dictyostelium discoideum. *Science (80-.).* **299**: 105–106.

Queller, D.C. & Strassmann, J.E. 2013. The veil of ignorance can favour biological cooperation. *Biol. Lett.* **9**: 1–4.

Queller, D.C., Zacchi, F., Cervo, R., Turillazzi, S., Henshaw, M.T., Santorelli, L. a, *et al.* 2000. Unrelated helpers in a social insect. *Nature* **405**: 784–787.

Quickfall, C.G. & Marshall, J.A.R. 2017. The evolution of mutualism with modifiers. *Ecol. Evol.* **7**: 6114–6118.

Rankin, D.J., Bargum, K. & Kokko, H. 2007. The tragedy of the commons in evolutionary biology. *Trends Ecol. Evol.* **22**: 643–651.

Rapoport, A. & Chammah, A.M. 1966. The Game of Chicken. *Am. Behav. Sci.* **10**: 10–28.

Ratnieks, F.L.W., Foster, K.R. & Wenseleers, T. 2006. Conflict Resolution in Insect Societies. *Annu. Rev. Entomol.* **51**: 581–608.

Ratnieks, F.L.W. & Reeve, H.K. 1992. Conflict in single-queen hymenopteran societies: the structure of conflict and processes that reduce conflict in advanced eusocial species. *J. Theor. Biol.* **158**: 33–65.

Rawls, J. 1971. *A Theory of Justice*. Harvard University Press.

Reeve, H.K. & Dugatkin, L.A. 1998. Why we need evolutionary game theory. In: *Game Theory and Animal Behavior*, pp. 304–311. Oxford University Press.

Richard, F., Glass, N.L. & Pringle, A. 2012. Cooperation among germinating spores facilitates the growth of the fungus, Neurospora crassa. *Biol. Lett.* **8**: 419–422.

Ridley, M. 2000. *Mendel's Demon: gene justice and the complexity of life*. Weidenfeld & Nicolson.

Ridley, M. & Grafen, A. 1981. Are green beard genes outlaws? *Anim. Behav.* **29**: 954–955.

Riedman, M.L. 1982. The Evolution of Alloparental Care and Adoption in Mammals and Birds. *Q. Riview Biol.* **57**: 405–435.

Riehl, C. 2013. Evolutionary routes to non-kin cooperative breeding in birds. *Proc. R. Soc. B Biol. Sci.* **280**: 20132245.

Rodrigues, A.M.M. & Gardner, A. 2016. The constant philopater hypothesis: A new life history invariant for dispersal evolution. *J. Evol. Biol.* **29**: 153–166.

Rothstein, S.I. & Barash, D.P. 1983. Gene conflicts and the concepts of outlaw and sheriff

alleles. *J. Soc. Biol. Syst.* **6**: 367–379.

Rousset, F. & Roze, D. 2007. Constraints on the Origin and Maintenance of Genetic Kin Recognition. *Evolution (N. Y).* **61**: 2320–2330.

Sachs, J.L., Mueller, U.G., Wilcox, T.P. & Bull, J.J. 2004. The Evolution of Cooperation. *Q. Rev. Biol.* **79**: 135–160.

Sandler, S. & Novitski, E. 1957. Meiotic drive as an evolutionary force. *Am. Nat.* **857**: 105–110.

Santema, P. & Clutton-Brock, T. 2013. Meerkat helpers increase sentinel behaviour and bipedal vigilance in the presence of pups. *Anim. Behav.* **85**: 655–661.

Santorelli, L.A., Kuspa, A., Shaulsky, G., Queller, D.C. & Strassmann, J.E. 2013. A new social gene in Dictyostelium discoideum, chtB. *BMC Evol. Biol.* **13**: 4. BMC Evolutionary Biology.

Santorelli, L.A., Thompson, C.R.L., Villegas, E., Svetz, J., Dinh, C., Parikh, A., *et al.* 2008. Facultative cheater mutants reveal the genetic complexity of cooperation in social amoebae. *Nature* **451**: 1107–1110.

Sasaki, T. & Uchida, S. 2014. Rewards and the evolution of cooperation in public good games. *Biol. Lett.* **10**: 20130903.

Scofield, V.L., Schlumpberger, J.M., West, L.A. & Weissman, I.L. 1982. Protochordate allorecognition is controlled by a MHC-like gene system. *Nature* **295**: 499–502.

Shaulsky, G. & Kessin, R.H. 2007. The Cold War of the social amoebae. *Curr. Biol.* **17**: 684–692.

Sigmund, K. 2010. *The Calculus of Selfishness*. Princeton University Press.

Sigmund, K. & Nowak, M.A. 1986. Evolutionary Game Theory. *Curr. Biol.* **9**: 503–505.

Silver, L.M. 1993. The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet.* **9**: 250–254.

Sinervo, B. & Lively, C.M. 1996. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* **380**: 240–243.

Skutch, A.F. 1935. Helpers at the nest. *Auk* **52**: 257–273.

Skyrms, B. 1996. *Evolution of the Social Contract*. Cambridge University Press.

Skyrms, B. 2004. *The Stag Hunt and the Evolution of Social Structure*. Cambridge University Press.

Smith, J., Queller, D.C. & Strassmann, J.E. 2014. Fruiting bodies of the social amoeba Dictyostelium discoideum increase spore transport by Drosophila. *BMC Evol. Biol.* **14**: 105.

Smukalla, S., Caldara, M., Pochet, N., Beauvais, A., Guadagnini, S., Yan, C., *et al.* 2008. FLO1 Is a Variable Green Beard Gene that Drives Biofilm-like Cooperation in Budding Yeast. *Cell* **135**: 726–737.

Springer, S.A., Crespi, B.J. & Swanson, W.J. 2011. Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture. *Mol. Ecol.* **20**: 2240–2257.

Strassmann, J.E., Gilbert, O.M. & Queller, D.C. 2011. Kin discrimination and cooperation in microbes. *Annu. Rev. Microbiol.* **65**: 349–367.

Strassmann, J.E. & Queller, D.C. 2012. Evolution of cooperation and control of cheating in a social microbe. *Proc. Natl. Acad. Sci.* **108**: 10855–10862.

Strassmann, J.E. & Queller, D.C. 2014. Privatization and property in biology. *Anim. Behav.* **92**: 305–311. Elsevier Ltd.

Strassmann, J.E., Zhu, Y. & Queller, D.C. 2000. Altruism and social cheating in the social amoeba Dictyostelium discoideum. *Nature* **408**: 965–967.

Sugden, R. 1986. *The Economics of Rights, Co-operation and Welfare*. Blackwell.

Svedberg, J., Hosseini, S., Chen, J., Vogan, A.A., Mozgova, I., Hennig, L., *et al.* 2018. Convergent evolution of complex genomic rearrangements in two fungal meiotic drive elements. *Nat. Commun.* **9**: 4242.

Taylor, P.D. 1992. Altruism in viscous populations - an inclusive fitness model. *Evol. Ecol.* **6**: 352–356.

Taylor, P.D. & Frank, S.A. 1996. How to make a kin selection model. *J. Theor. Biol.* **180**: 27–37.

Taylor, P.D., Wild, G. & Gardner, A. 2007. Direct fitness or inclusive fitness: how shall we model kin selection? *J. Evol. Biol.* **20**: 301–309.

Thompson, C.R.L. & Kay, R.R. 2000. The role of DIF-1 signaling in Dictyostelium development. *Mol. Cell* **6**: 1509–1514.

Thompson, M.J. & Jiggins, C.D. 2014. Supergenes and their role in evolution. *Heredity (Edinb).* **113**: 1–8.

Thomson, G.J. & Feldman, M.W. 1976. Population genetics of modifiers of meiotic drive III. Equilibrium analysis of a general model for the genetic control of segregation distortion. *Theor. Popul. Biol.* **10**: 10–25.

Travisano, M. & Velicer, G.J. 2004. Strategies of microbial cheater control. *Trends Microbiol.* **12**: 72–78.

Trible, W. & Ross, K.G. 2015. Chemical communication of queen supergene status in an ant. *J. Evol. Biol.* **29**: 502–513.

Trivers, R.L. 1971. The Evolution of Reciprocal Altruism. *Q. Rev. Biol.* **46**: 35–57.

Turner, P.E. & Chao, L. 1999. Prisoner's dilemma in an RNA virus. *Nature* **398**: 441–443.

Unckless, R.L., Clark, A.G. & Messer, P.W. 2017. Evolution of resistance against CRISPR/Cas9 gene drive. *Genetics* **205**: 827–841.

Unterweger, D. & Griffin, A.S. 2016. Nice or nasty: Protein translocation between bacteria and the different forms of response. *Proc. Natl. Acad. Sci.* **113**: 8559–8561.

Van Dyken, J.D., Linksvayer, T.A. & Wade, M.J. 2011. Kin Selection–Mutation Balance: A Model for the Origin, Maintenance, and Consequences of Social Cheating. *Am. Nat.* **177**: 288–300.

Vassallo, C.N., Cao, P., Conklin, A., Finkelstein, H., Hayes, C.S. & Wall, D. 2017. Infectious polymorphic toxins delivered by outer membrane exchange discriminate kin in myxobacteria. *Elife* **6**: e29397.

Velicer, G.J. 2003. Social strife in the microbial world. *Trends Microbiol.* **11**: 330–337.

Velicer, G.J., Kroos, L. & Lenski, R.E. 2000. Developmental cheating in the social bacterium Myxococcus xanthus. *Nature* **404**: 598–601.

Verstrepen, K.J., Jansen, A., Lewitter, F. & Fink, G.R. 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* **37**: 986–990.

Wade, M.J. & Beeman, R.W. 1994. The population dynamics of maternal-effect selfish genes. *Genetics* **138**: 1309–1314.

Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., *et al.* 2013. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**: 664–8.

Wang, Y. & Shaulsky, G. 2015. TgrC1 Has Distinct Functions in Dictyostelium Development and Allorecognition. *PLoS One* **10**: e0124270.

Warr, P.G. 1983. The private provision of a public good is independent of the distribution of income. *Econ. Lett.* **13**: 207–211.

Wechsler, T., Kümmerli, R. & Dobay, A. 2019. Understanding policing as a mechanism of cheater control in cooperating bacteria. *J. Evol. Biol.* **32**: 412–424.

Wenseleers, T., Helanterä, H., Hart, A. & Ratnieks, F.L.W. 2004. Worker reproduction and policing in insect societies: An ESS analysis. *J. Evol. Biol.* **17**: 1035–1047.

Wenseleers, T. & Ratnieks, F.L.W. 2006. Comparative Analysis of Worker Reproduction and Policing in Eusocial Hymenoptera Supports Relatedness Theory. *Am. Nat.* **168**: E163–E179.

Wenseleers, T. & Ratnieks, F.L.W. 2004. Tragedy of the commons in Melipona bees. *Proc. Biol. Sci.* **271 Suppl**: S310-2.

Werren, J.H., Nur, U. & Wu, C.I. 1988. Selfish genetic elements. *Trends Ecol. Evol.* **3**: 297–302.

West, S.A. & Buckling, A. 2003. Cooperation, virulence and siderophore production in bacterial parasites. *Proc. Biol. Sci.* **270**: 37–44.

West, S.A., El Mouden, C. & Gardner, A. 2011. Sixteen common misconceptions about the evolution of cooperation in humans. *Evol. Hum. Behav.* **32**: 231–262.

West, S.A. & Gardner, A. 2013. Adaptation and inclusive fitness. *Curr. Biol.* **23**: R577-584.

West, S.A., Griffin, A.S. & Gardner, A. 2007a. Evolutionary explanations for cooperation. *Curr. Biol.* **17**: R661–R672.

West, S.A., Griffin, A.S. & Gardner, A. 2007b. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evol. Biol.* **20**: 415–432.

West, S.A., Griffin, A.S., Gardner, A. & Diggle, S.P. 2006. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **4**: 597–607.

West, S.A., Toby Kiers, E., Pen, I. & Denison, R.F. 2002. Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15**: 830–837.

Wielgoss, S., Fiegna, F., Rendueles, O., Yu, Y.T.N. & Velicer, G.J. 2018. Kin discrimination and outer membrane exchange in Myxococcus xanthus: A comparative analysis among natural isolates. *Mol. Ecol.* **27**: 3146–3158.

Williams, G.C. 1966. *Adaptation and Natural Selection: a critique of some evolutionary thought*. Princeton University Press.

Williams, G.C. & Williams, D.C. 1957. Natural Selection of Individually Harmful Social Adaptations Among Sibs With Special Reference to Social Insects. *Evolution (N. Y).* **11**: 32.

Wilson, D.S. 1975a. A theory of group selection. *Proc. Natl. Acad. Sci.* **72**: 143–146.

Wilson, D.S. 1977. Structured demes and the evolution of group-advantageous traits. *Am. Nat.* **111**: 157–185.

Wilson, E.O. 1975b. *Sociobiology: The New Synthesis*. Harvard University Press.

Wilson, E.O. & Hölldobler, B. 1990. *The Ants*. Harvard University Press.

Wolf, J.B., Howie, J.A., Parkinson, K., Gruenheit, N., Melo, D., Rozen, D., *et al.* 2015. Fitness trade-offs result in the illusion of social success. *Curr. Biol.* **25**: 1086–1090.

Wright, S. 1945. Tempo and Mode in Evolution : A Critical Review. *Ecology* **26**: 415–419.

Xavier, J.B., Kim, W. & Foster, K.R. 2011. A molecular mechanism that stabilizes cooperative secretions in Pseudomonas aeruginosa. *Mol. Microbiol.* **79**: 166–179.

Zhang, H. & Chen, S. 2016. Tag-mediated cooperation with non-deterministic genotype-phenotype mapping. *Europhys. Lett.* **113**: 28008.