Life in space

Samuel R. Levin

Hertford College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2019

Abstract

For nearly half a century, inclusive fitness theory has formed a bedrock of whole-organism biology. It has helped explain social behaviours across the tree of life, including, via the major transitions view of evolution, the history of life itself. However, some significant questions remain. In this thesis, I consider challenges at the frontiers of inclusive fitness, in time, theory, and space. Specifically: (i) I study the evolution of cooperation early in the history of life. I develop models of simple molecular replicators, resolving previously puzzling results and identifying important life-history features of replicators for cooperation. (ii) I use kin selection models to develop a framework for understanding RNA cooperation and guiding empirical work on the origins of life. (iii) I develop and test a new hypothesis for the origin of the genome, the first major transition in individuality. (iv) I address criticisms of inclusive fitness theory, developing conceptual arguments for why it is a useful tool despite its flaws. (v) I show that recent mathematical criticisms of inclusive fitness misapplied inclusive fitness, and extend their models, recovering inclusive fitness maximisation. I provide formal arguments for a broader range of the theory's application than previously thought by some mathematical biologists. (vi) I show that a recent paper about the effects of divorce on honest signalling in birds miscalculated inclusive fitness, and develop a formal model that predicts an effect of inclusive fitness on signalling in birds, which is supported by the data. (vii) Finally, I consider the universality of natural selection and major transitions in individuality, demonstrating how evolutionary theory can be used as a powerful tool in astrobiology. As a whole, this body of work extends the range of inclusive fitness theory's application earlier in evolutionary time, to a wider range of biological scenarios, and further in space.

Life in space

Samuel R. Levin
Hertford College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Hilary 2019

Declaration

I declare that this thesis was composed by myself and that the work contained herein is my own except where explicitly stated in the text. The work has not been submitted for any degree or professional qualification except as specified.

Samuel R. Levin

Acknowledgements

There are a handful of people whose comments, discussions, feedback, and guidance contributed greatly to this thesis and to my development as a scientist. They are: Jay Biernaskie, Kevin Foster, Ashleigh Griffin, Rafe Kennedy, Dave Queller, Joan Strassmann, and Geoff Wild. Thank you. Long before this thesis began, Kathy Erickson introduced me to maths, which is why this biology thesis has so many equations.

My labmates, Anna, Becca, Chucky, David, Daniel, JB, Josh, Max, Mel, Philip, Shana, and Steffi, made implausible locales like E7, Magdalen, NRH, and finally the porta-cabins a great home for science. A special thanks is due to the members of SUMS Club: Asher, Gijsbert, Guy, Mati, Miguel, and Tom. This thesis would not have been possible without your barrage of insights, arguments, and, of course, sums.

Alan, despite not supervising me in any official capacity, took the time to teach me about natural selection, and about thinking clearly. I am deeply grateful for your patience and wisdom.

I am particularly thankful to my grandmother, Tinka, for her endless support and encouragement, and to Scottie, for introducing me to the joys of weebeasties, bloodroot, and the rest of the natural world. And, of course, Beth, for providing a haven for so many years.

Finally, I thank my supervisor, Stu, who possesses the rare combination of being a brilliant scientist, communicator, and mentor, and is unflinchingly generous with all three. Thank you for a life-time's worth of knowledge.

This thesis is dedicated to my parents. To my mom, for teaching me what it meant to be a scientist. And to my dad, for showing me what it meant to have resolve.

Publications and Contributions

The following published papers have arisen from this thesis, and are presented in Chapters 2, 3, 8 and the Appendix.

- **Chapter 2**
  - **Levin, S.R.** & West, S.A. (2017) The evolution of cooperation in simple molecular replicators. *Proceedings of the Royal Society London Series B: Biological Sciences* 284 (1864), 20171967
    - I conceived of the manuscript with feedback from SAW. I carried out the modelling and wrote the first draft of the manuscript. SAW and I contributed equally to the preparation of the manuscript in its final form.
- **Chapter 3**
  - **Levin, S.R.** & West, S.A. (2017) Kin selection in the RNA world. *Life* 7 (4), 53
    - I conceived of the manuscript with feedback from SAW. I carried out the modelling and wrote the first draft of the manuscript. SAW and I contributed equally to the preparation of the manuscript in its final form.
- **Chapter 8**
  - **Levin, S.R.**, Scott, T.W., Cooper, H.S. & West, S.A. (2017) Darwin's aliens. *International Journal of Astrobiology* 18 (1), 1-9
    - SAW and I conceived of the manuscript. I wrote the first draft, and TWS, SAW and I contributed equally to preparing the final draft. HSC created the illustrations. I created Figure 3 and TS created Figure 1.
- **Appendix A**
  - Cooper, G.A.*, **Levin, S.R.***, Wild, G. & West, S.A. (2018) Modelling relatedness and demography in social evolution. *Evolution Letters* 2 (4), 260-271. *These authors are joint first authors.

- All authors conceived of the manuscript and contributed equally to the writing of the manuscript. I carried out the modelling for Box 1 and created Figure 1.

The following manuscripts have arisen from this thesis, and are presented in Chapters 4, 5, 6, and 7.

- **Chapter 4**
  - **Levin, S.R.**, Gandon, S., & West, S.A. The social coevolution hypothesis for the origin of the genome. *In preparation*
    - I conceived of the manuscript with feedback from SAW. I carried out the modelling with help from SG and feedback from SAW. I wrote the first draft. All authors contributed equally to the preparation of the final manuscript.
- **Chapter 5**
  - **Levin, S.R.** & Grafen, A. (*In press*) Inclusive fitness is an indispensable approximation for understanding organismal design. *Evolution*
    - I conceived of the manuscript with feedback from AG and wrote the first draft. All authors contributed equally to the preparation of the manuscript in its final form.
- **Chapter 6**
  - **Levin, S.R.** & Grafen, A. Extending the range of additivity in using inclusive fitness. *Under review at Evolution Letters*
    - I conceived of the manuscript, wrote the first draft, and carried out the modelling, with feedback from AG. All authors contributed equally the preparation of the manuscript in its final form.
- **Chapter 7**
  - **Levin, S.R.**, Caro, S.M., Griffin, A.S., and West, S.A. Honest signalling and the double counting of inclusive fitness. *Under review at Evolution Letters*
    - I conceived of the manuscript and carried out the modelling. SMC carried out the data analysis. All authors contributed equally to the preparation of the manuscript.

Abstract

For nearly half a century, inclusive fitness theory has formed a bedrock of whole-organism biology. It has helped explain social behaviours across the tree of life, including, via the major transitions view of evolution, the history of life itself. However, some significant questions remain. In this thesis, I consider challenges at the frontiers of inclusive fitness, in time, theory, and space. Specifically: (i) I study the evolution of cooperation early in the history of life. I develop models of simple molecular replicators, resolving previously puzzling results and identifying important life-history features of replicators for cooperation. (ii) I use kin selection models to develop a framework for understanding RNA cooperation and guiding empirical work on the origins of life. (iii) I develop and test a new hypothesis for the origin of the genome, the first major transition in individuality. (iv) I address criticisms of inclusive fitness theory, developing conceptual arguments for why it is a useful tool despite its flaws. (v) I show that recent mathematical criticisms of inclusive fitness misapplied inclusive fitness, and extend their models, recovering inclusive fitness maximisation. I provide formal arguments for a broader range of the theory's application than previously thought by some mathematical biologists. (vi) I show that a recent paper about the effects of divorce on honest signalling in birds miscalculated inclusive fitness, and develop a formal model that predicts an effect of inclusive fitness on signalling in birds, which is supported by the data. (vii) Finally, I consider the universality of natural selection and major transitions in individuality, demonstrating how evolutionary theory can be used as a powerful tool in astrobiology. As a whole, this body of work extends the range of inclusive fitness theory's application earlier in evolutionary time, to a wider range of biological scenarios, and further in space.

Contents

1 Introduction

Organismal purpose

Organisms appear designed as if for a purpose. In the absence of something or someone having designed them that way, this presents a puzzle. Darwin (1859) largely resolved this puzzle with his theory of natural selection. Natural selection explains both the process by which organisms acquire the appearance of design and the purpose for which they appear designed. The process is the inheritance of variation linked to differential reproductive success, and the purpose is the maximisation of reproductive success (Darwin, 1859; Fisher, 1930; Grafen, 2014; Gardner, 2017). While Darwin's arguments were verbal, Fisher (1930) later provided formal support, showing that mean fitness increases due to the action of natural selection (Price, 1972; Frank, 2012a; Grafen, 2015; Queller, 2017).

But this theory is at odds with the numerous examples of traits that do not maximise reproductive success. The most notable example is altruism, in which an organism reduces its own number of offspring in order to increase the number of offspring of others. Hamilton (1964) explained that this failure arises because Darwin's original theory does not take into account social interactions between individuals. An individual's number of offspring is affected not just by its own actions, but the actions of others, and a proper measure of fitness must account for this.

More precisely, an organism's mean offspring number, which determines the direction of selection, includes the sum of the expected effects on the individual's offspring number of all individuals in the population, including itself. Hamilton termed this expanded notion of fitness 'neighbour modulated fitness', and pointed out that such a measure is 'unwieldy' (Hamilton, 1964). This is partly because, in practice, it requires knowing the relationship between genotype and phenotype as well as the actual genotypes of individuals, two things we rarely know (Hamilton, 1964; Grafen, 1982, 1984; Queller, 1996).

Inclusive fitness

Hamilton (1964) pointed out that a simpler alternative is to take the perspective of a focal individual, and sum the fitness effects the individual has on others. Hamilton termed this alternative metric 'inclusive fitness'. More precisely, inclusive fitness measures an individual's adult offspring number, stripped of all social effects, and a weighted sum of the effects the individual has on all individuals in the population (including itself). The weightings are degrees of relatedness, where relatedness (R) is a measure of genetic similarity (with $R = 0$ for a random member of the population, and $R = 1$ for a clone) (Hamilton, 1964, 1970; Grafen, 1985). Following Fisher (1930), Hamilton showed, under some assumptions, that inclusive fitness increases due to the action of natural selection (Hamilton, 1964, 1970). Accordingly, we can understand the purpose for which organisms appear designed as being to maximise their inclusive fitness (Hamilton, 1964, 1970; Grafen, 1984; Queller, 1992; Frank, 1998; Grafen, 2006; Gardner et al., 2011; West and Gardner, 2013; Foster, 2009; Rousset, 2015; Lehmann et al., 2016; Taylor, 2017).

This provided an explanation for previously aberrant behaviours such as altruism, and sparked an entire field of empirical social biology (Krebs and Davies, 1978, 1987; Foster, 2009; Krebs and Davies, 2009; Westneat and Fox, 2010; Davies et al., 2012). Hamilton's logic, which underpins this field, can be captured in a simple inequality, known as 'Hamilton's rule'. Given a trait that confers a fitness benefit, B, to some recipients, at a cost, C, to an actor, with relatedness between the recipients and

the actor, R, Hamilton's rule says that such a trait will increase in frequency if $RB - C > 0$ (Hamilton, 1964, 1970; Queller, 1992; Gardner et al., 2011). Altruism can be explained if the weighted benefits to relatives outweigh the costs to self.

Further, the logic is not limited to explaining altruism. We can conceptualise all social behaviours according to their relatedness-weighted effects. This provides a framework for understanding classic selfishness (negative effect on others, positive effect on self), but also mutualism (positive effect on both) and even spite (negative effect on both) (Hamilton, 1964, 1970). More generally, we can expect organisms, at equilibrium, to behave as though they value the effects of their behaviours according to their relatednesses to the individuals affected by them.

While Hamilton's (1964) original model was applied to absolute effects on offspring number, extensions have since generalised the approach (Queller, 1985; Gardner et al., 2011). Queller (1992) derived a statistical version of Hamilton's rule, in which the terms are regressions on phenotypes. This form allows a causal explanation of behaviour for any type of trait or population structure, and can be considered as a generalisation of Darwin's theory to incorporate social interactions (Queller, 1992; Gardner et al., 2011).

Major transitions

Further, social behaviours are not limited to birds helping at the nest, or bees pollinating a flower. The very structure of an organism depends on extreme cooperation between its parts. Genes cooperate to form functioning genomes, mitochondria and nuclei cooperate to form eukaryotic cells, cells cooperate to form multicellular bodies, and, in some cases, multicellular bodies cooperate to form eusocial societies. The history of life has been shaped by the cooperation of these entities, where, along the way, previously independent units collaborated to form new, cohesively functioning wholes (Smith and Szathmáry, 1995; Bourke, 2011a; West et al., 2015).

These events are known as major transitions in individuality, and they punctuate the rise in complexity on Earth (Smith and Szathmáry, 1995; Queller, 1997; Bourke,

2011a; West et al., 2015). While events like mutations and duplications can also cause an increase in complexity, these increases tend to be gradual and reversible. On the other hand, major transitions in individuality tend to be large and irreversible (Queller, 1997; Bourke, 2011a; West et al., 2015). For example, before multicellularity, all life consisted of at most one cell working to maximise its inclusive fitness. Afterwards, life included organisms containing a near-unlimited number of cells, specialising on different tasks to maximise the inclusive fitness of the entire collection.

For such complexity to be maintained, it has to be possible for adaptations to accrue at the level of the collection (Queller, 1997; Gardner and Grafen, 2009; West et al., 2015). This can only be true if selection at the lower levels does not disrupt selection at the higher levels. Consider, for example, the evolution of multicellularity. It requires the somatic cells giving up reproduction altogether in order that the the germ cells can reproduce into the next generation. This is an extreme form of cooperation, and all else being equal, we might expect the somatic cells to be selected to reproduce themselves, disrupting the organism.

Such extreme cooperation can be explained, in part, if there is near complete alignment of interests between the adjoining units, such that there is negligible conflict within the larger collection (West et al., 2015; Gardner and Grafen, 2009; Strassmann and Queller, 2010; Bourke, 2011a). In the case of multicellular organisms, this arises, because (among other things) they start their life cycle as a single-celled zygote. As a result, all cells within the organism are genetic clones, and have a relatedness of $R = 1$. Along with the early sequestration of the germ-line, this effectively eliminates conflict between the cells, aligning their interests and allowing adaptations to accrue at the level of the collection of cells. Of course, these adaptations will only be favoured if there is also some benefit to forming a group, for example defending against predators (Bourke, 2011a; West et al., 2015; Queller and Strassmann, 1998; Koschwanez et al., 2013; Fisher et al., 2017, 2016; Kapsetaki et al., 2016), but the alignment paves the way for such adaptations.

More generally, complete alignment of interests is permitted by maximal relatedness (West et al., 2015; Gardner and Grafen, 2009). While some major transitions,

like the origin of multicellularity, have occurred between individuals within a species, and others, like the origin of the eukaryotic cell, have occurred between-species, they all seem to be driven by a similar alignment of interests, which can be conceptualised as maximal relatedness (Queller, 1997; Foster and Wenseleers, 2006; West et al., 2015; Gardner and Grafen, 2009; Frank, 2013; Boomsma and Gawne, 2018). Inclusive fitness theory then, as the expanded view of natural selection, explains not just social behaviours, but the very structure and history of life itself.

Open questions

However, questions remain. First, while later transitions, such as multicellularity and eusociality, have been well studied both empirically and theoretically, the earliest transition, the origin of the genome, has been relatively under-explored (Smith and Szathmáry, 1995; Levin and West, 2017b). Can we understand the coming together of independent replicators to form a genome as being explained by the same forces that drive later transitions? What specific features of simple replicating molecules would have favoured the evolution of cooperation? While replicators of the same type, or species, might have cooperated, ultimately the genome evolved as a collection of different types of replicators, each maintaining their own reproduction. What might have favoured such between-type cooperation?

Second, inclusive fitness theory, despite its empirical successes, has been controversial for decades (for example, see Nowak et al. (2010) and replies, e.g. Abbot et al. (2011), Bourke (2011b), and Queller (2016)). Specifically, inclusive fitness has long been criticised for its assumptions, most notably that of additivity of fitness effects (Cavalli-Sforza and Feldman, 1978; Uyenoyama and Feldman, 1982; Karlin and Matessi, 1983; Queller, 1985; Allen et al., 2013). Hamilton's (1964) original proof assumed that the fitness effects of an actor on a recipient combine linearly with its existing offspring number (and therefore, implicitly, with the effects of others). This has led a number of authors to suggest abandoning inclusive fitness altogether (e.g. Nowak et al., 2010; Nowak and Allen, 2015; Nowak et al., 2017; Allen et al., 2013; Allen and Nowak, 2016; Allen, 2015), with some suggesting a

return to Hamilton's 'unwieldy' neighbour modulated fitness, or mean offspring number (Okasha and Martens, 2016; Lehmann et al., 2015; Allen and Nowak, 2015). In the extreme, recent models claim to show formally that inclusive fitness is not maximised by natural selection under non-additivity (Lehmann et al., 2015; Okasha and Martens, 2016). What are biologists, who have been using inclusive fitness for decades, to make of these mathematical arguments?

And finally, to what degree is inclusive fitness theory universal? Inclusive fitness arguments were developed with domestic organisms in mind. However, astrobiology, the study of life on other planets, is a rapidly growing field, with large amounts of resources being dedicated towards detecting and searching for life on other planets (Des Marais et al., 2008; Horneck et al., 2016). The generality of evolutionary theory might provide a powerful tool in such a search, because it has the potential to make phenotype-level predictions, independent of chemical details (Levin et al., 2017). To what degree do we expect natural selection to hold on other planets? Can we expect the rise in complexity in space to be driven by a similar form of major transition in individuality?

Thesis outline

As with any well-developed field, the frontiers tend to lie at the edges. In this thesis, I consider some of the frontiers of the field, in time (the origin of the genome), theory (inclusive fitness under non-additivity), and space (the universality of inclusive fitness). Specifically:

In **Chapter 2**, I study the evolution of cooperation in simple molecular replicators, a possible first step on the path towards genomes. Previous models used simulations to argue that limited diffusion on surfaces could explain such cooperation (Boerlijst and Hogeweg, 1991, 1995; Cronhjort and Blomberg, 1997; Szabó et al., 2002; Sardanyés and Solé, 2007; Bianconi et al., 2013; Shay et al., 2015; McCaskill et al., 2001). However, a standard result from social evolution theory is that limited diffusion is not sufficient to favour cooperation (Taylor, 1992). I develop social evolution

models to study the effects of limited diffusion on cooperation in replicators. I show that: (i) replicators can be considered to be cooperating as a result of kin selection; (ii) limited diffusion on its own does not favour cooperation; and (iii) the addition of overlapping generations, a likely trait of molecular replicators, promotes cooperation.

In **Chapter 3**, I extend the work of Chapter 2, to consider the role of RNA biology in cooperation between replicators more generally. Various steps in the RNA world required cooperation (Higgs and Lehman, 2015; Levin and West, 2017a). I develop a very simple model of RNA cooperation and then elaborate it to study three relevant features of RNA biology. I show: (i) that RNAs are likely to express partial cooperation; (ii) that RNAs will need mechanisms for overcoming local competition; and (iii) in a specific example of RNA cooperation, persistence after replication and limited offspring diffusion allow for cooperation to overcome competition. More generally, I show how kin selection can unify previously disparate answers to the question of RNA-world cooperation.

In **Chapter 4**, I move beyond cooperation between replicators of the same type, to consider why cooperation between different types of replicators, required for the genome, might have evolved. Existing hypotheses to solve the problem of replicator cooperation require restrictive assumptions, such as the evolution of a cell membrane before the evolution of a genome, or very particular patterns of diffusion on special types of surfaces (Smith and Szathmáry, 1995; Frank, 1994; Shay et al., 2015). I develop an alternative hypothesis to explain such cooperation. I show that the tendency to physically associate to others and cooperative enzymatic activity can coevolve, leading to the evolution of physically linked cooperative replicators, akin to a primitive genome.

In **Chapter 5**, I turn to the problem of inclusive fitness maximisation more generally. Inclusive fitness has long been criticised for its assumptions. For decades authors have pointed out that mean offspring number does a better job at predicting

gene frequency change in a wider range of scenarios. In this chapter I present conceptual arguments for why inclusive fitness, despite its drawbacks, is a more suitable maximand for biologists, illustrating one of the key points with a simple model. I discuss the empirical relevance of key assumptions, and highlight the scenarios that could cause biologists to suspect inclusive fitness failure.

In **Chapter 6**, I address some of the technical challenges to inclusive fitness maximisation. Several recent papers have claimed to show the failure of inclusive fitness maximisation under non-additivity of fitness effects (Lehmann et al., 2015; Okasha and Martens, 2016). I extend these models, showing (i) that the authors failed to consider the correct measure of inclusive fitness, as defined by Hamilton (1964); (ii) how to capture inclusive fitness correctly in two specific mathematical scenarios; and (iii) that under biologically realistic assumptions, inclusive fitness maximisation is indeed recovered in these models.

In **Chapter 7**, I consider an example in which researchers have miscalculated inclusive fitness, leading to erroneous predictions. Bebbington and Kingma (2017) predict that begging chicks, in adjusting their signal honesty to maximise inclusive fitness, should be indifferent to whether their parents divorce. However, they commit an error known as double-counting. I develop formal models to show that, under standard assumptions, divorce does matter. I discuss the existing data, and argue that it supports the case that chicks should be less honest when their parents divorce. I discuss the importance of using the correct inclusive fitness, and of making verbal arguments formal to avoid errors.

In **Chapter 8**, I consider the universality of inclusive fitness arguments. What can evolutionary theory tell us about alien life forms? Previous work has extrapolated from examples on Earth, which potentially limits the arguments to a sample size of one. I develop an argument for why aliens must undergo natural selection, and consider what this tells us, from a theoretical perspective, about extra-terrestrials.

Further, I argue that we might expect complex aliens to be the product of major transitions, and suggests that this allows us to make Earth-independent predictions about their make-up. I consider the role of evolutionary theory in the future of astrobiological research.

Finally, in **Chapter 9**, I summarise the main results of Chapters 2-8. I discuss broader implications for evolutionary theory, and future directions for origins of life, inclusive fitness maximisation, and astrobiology research.

I have not provided a more extensive formal literature review in this introduction, as I review the relevant literature in Chapters 2-8.

2

The evolution of cooperation in simple molecular replicators

PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Levin SR, West SA. 2017 The evolution of cooperation in simple molecular replicators. *Proc. R. Soc. B* **284**: 20171967. http://dx.doi.org/10.1098/rspb.2017.1967

Received: 1 September 2017
Accepted: 5 September 2017

Subject Category:
Evolution

Subject Areas:
evolution, theoretical biology

Keywords:
molecular replicators, cooperation, kin selection, social evolution, limited diffusion, genome origin

Author for correspondence:
Samuel R. Levin
e-mail: samuel.levin@zoo.ox.ac.uk

THE ROYAL SOCIETY
PUBLISHING

The evolution of cooperation in simple molecular replicators

Samuel R. Levin and Stuart A. West

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

SRL, 0000-0002-9588-7729

In order for the first genomes to evolve, independent replicators had to act cooperatively, with some reducing their own replication rate to help copy others. It has been argued that limited diffusion explains this early cooperation. However, social evolution models have shown that limited diffusion on its own often does not favour cooperation. Here we model early replicators using social evolution tools. We show that: (i) replicators can be considered to be cooperating as a result of kin selection; (ii) limited diffusion on its own does not favour cooperation; and (iii) the addition of overlapping generations, probably a general trait of molecular replicators, promotes cooperation. These results suggest key life-history features in the evolution of the genome and that the same factors can favour cooperation across the entire tree of life.

1. Introduction

Genomes are made up of genes, or replicators, which work together to produce an organism. These genes specialize in different tasks—for example, some produce replication machinery to copy the other genes, while others focus on acquiring energy for this process. However, life began with independent replicators, whose sole purpose was to copy themselves [1–3]. Thus, at some point, before the last universal common ancestor, independent replicators came together to form a rudimentary genome. At least some of these replicators had to focus on the task of replicating others, reducing their own replication rate in the process. Thus, to get from the first replicators to the first genome, independent replicators must have acted cooperatively [2–4]. This poses a problem, because we expect natural selection to favour individuals that selfishly replicate themselves. For example, imagine a mutant replicator that, rather than help copy others, replicated itself instead, and therefore had a higher replication rate than its neighbours. All else being equal, this mutant should prevail. So why would early replicators cooperate?

It has been argued that limited diffusion or dispersal could explain cooperation between early replicators [5–12]. The existence of early replicators on surfaces, such as rocks, would have led to relatively limited diffusion [3,13]. A number of simulation studies have examined this possibility, assuming that replicators exist on connected nodes in a two dimensional lattice. In these lattice models, when replicators replicate, their offspring can move one, two, or many nodes away. How far offspring can travel in a given step determines whether the system has limited diffusion or is well mixed. These simulations have shown that limited diffusion can favour the evolution of cooperative replicators, who help others replicate [5–12]. Limited diffusion keeps cooperators together, and so their cooperation is directed towards other cooperators, allowing them to outcompete non-cooperators.

However, this suggested role of limited diffusion raises two further questions. First, limited dispersal has previously been argued to explain cooperation in organisms ranging from bacteria to birds because it keeps relatives together [14–20]. In these cases, cooperation is favoured because it is directed at relatives who share the same genes, termed kin selection [14]. Can we think of this early cooperation between replicators as being favoured by kin selection, analogous to that in higher organisms? If so, we could make broad generalizations about

the factors that have favoured cooperation, across different biological levels, as life on earth evolved.

Second, theoretical kin selection models have shown that limited dispersal, on its own, does not favour cooperation (reviewed by [21,22]). Although limited dispersal increases the likelihood that cooperation can be directed towards relatives, it also increases competition between relatives. Taylor [23,24] showed that in the simplest case, these effects exactly cancel, and that the rate of dispersal does not influence selection for cooperation. Since then, a number of models have shown that limited dispersal can favour cooperation, but only if additional factors are added, such as overlapping generations, or dispersal in groups (buds), which allow the benefits of increased relatedness to outweigh the extra competition (e.g. [25–27]). Consequently, we must ask how limited diffusion manages to favour cooperation in these replicator models.

We develop theoretical models to address these two questions. We focus on a specific example of replicator biology, termed the trans-replicase system, because it is one of the simplest forms of molecular cooperation. First, we develop a simple kin selection model to examine whether we can think of limited diffusion as favouring cooperation in trans-replicases by kin selection [28]. This model allows us to compare the evolution of cooperation in a simple replicator with models developed to explain cooperation across a range of other taxa. Second, we develop a more spatially explicit model, to ask how limited diffusion might favour cooperation among replicators. We develop a relatively simple model to capture the key features of the previous simulations, but where we can obtain an analytical solution [5–12].

2. Heuristic overview

We start by developing the simplest possible, heuristic model. The purpose of this is to try to capture the biology of an early replicator using the tools of social evolution theory [29–31]. This kind of streamlined model aims to cut out all but the most essential biological and biochemical details, to capture a wider range of possible biologies, and identify key parameters. We sacrifice realism for generality and insight.

We model one possible system for cooperation in replicators: the trans-acting replicase or trans-replicase (figure 1) [12]. There are a variety of possible biologies for early replicators, but the trans-replicase model is one of the simplest. Trans-replicases are replicating molecules that, upon replicating, through mechanisms such as alternate folding, can express one of two phenotypes: (i) replicases, which are molecules that can copy other replicators, or (ii) templates, which can be copied by a replicase but do not act as replicases (figure 1). The replicase phenotype can be considered to be cooperative, because it reduces its own replication rate in order to increase the replication rate of others, and the template phenotype to be relatively selfish. An individual maintains its phenotype of being either a replicase or a template for life, such that any given individual is either a replicase or a template.

We assume that when a replicator is copied, the new copy (offspring) folds to become a replicase with probability x and a template with probability $1-x$. Mutations could cause individuals to express the cooperative replicase phenotype with higher or lower probability—creating more or less cooperative strategies. We are envisaging phenotypic variation

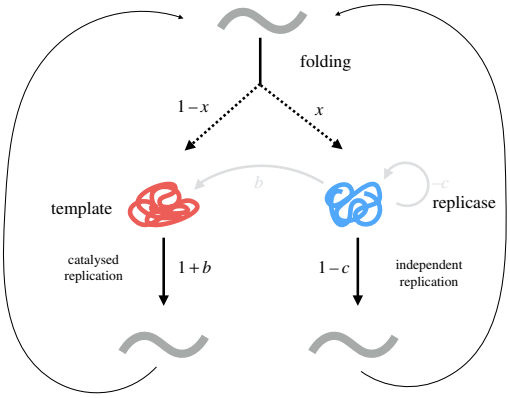

Figure 1. Life cycle for a trans-replicase. Individuals replicate and can express one of two phenotypes (e.g. through alternative folding patterns). With probability x they become a replicase, otherwise, with probability $1-x$, they become a template. All individuals have a baseline replication rate of 1. Replicases incur a cost, c, in terms of replication rate, and enzymatically increase the replication rate of templates by a factor of b. Replicases replicate at a rate of $1-c$, and templates replicate at a rate of 1, unless they interact with a replicase, in which case they replicate at rate $1+b$.

generated by alternate folding, but our model captures other ways to generate variable phenotypes. Although a trans-replicase capable of self-sustaining in a pre-biotic world has yet to be identified, the development of simple RNA molecules capable of template directed synthesis suggest their plausibility [32–37].

Individual replicators have a baseline replication rate, or fitness, of 1. A replicase experiences a replication rate reduced by c, which can be considered the cost of helping or cooperating with templates. If replicases cannot replicate, then $c=1$. The presence of replicases increase the replication rate of a template by a factor of by, where y is the average proportion of replicators which are replicases over the scale at which replicators can interact (the social group). This increase in the template replication rate can be considered the benefit of the cooperative or helpful act. When there are a higher fraction of replicases in the social group (y), there is a greater likelihood of any template being helped. Replicases do not catalyse the replication of other replicases, and so do not provide a benefit to replicases.

We can write the expected fitness of a focal individual (w) as the summed fitness of its replicase ($1-c$) and template ($1+by$) offspring, multiplied by their relative frequency, which is x and $1-x$, respectively, giving

$$w(x,y) = x(1-c) + (1-x)(1+by). \tag{2.1}$$

We are considering an asexual population, and so ignoring mutation, the strategy or phenotype of a replicator and the copies (offspring) that it produces will be the same, x. Consequently, equation (2.1) can also be conceptualized as the sum of the likelihood that a replicator developed as a replicase multiplied by its fitness in that scenario, and the likelihood that a replicator developed as a template multiplied by its fitness in that scenario. More traditionally, equation (2.1) is conceptualized as the average reproductive value of the focal offspring [38].

We seek the evolutionary stable strategy (ESS), x^*, which cannot be beaten by any other strategy [39]. Taylor [29] and Frank [31] developed an approach for determining the ESS in social models. Assuming selection is weak, candidate ESSs occur where the derivative, with respect to deviations in x (also known as the 'inclusive fitness effect') equals zero:

$$\frac{\mathrm{d}w}{\mathrm{d}x} = \frac{\partial w}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}x} + \frac{\partial w}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}x} = 0. \tag{2.2}$$

The $\mathrm{d}y/\mathrm{d}x$ term is the slope of the regression of a random partner's phenotype on the focal individual's, and can be replaced with r [29], the standard coefficient of relatedness [40–42]. Relatedness, r, is a measure of genetic similarity between our focal individual and the other individuals on the patch. In our model, r is the likelihood that our focal individual shares the same gene at a given locus with another individual on their patch, relative to a random member of the population. Relatedness can arise for a number of reasons, and r represents a summary of all details about the population structure. This kind of approach has proved useful for linking data with theory, because a simple model can then be applied to a variety of different biological cases, where a positive relatedness arises for different reasons [43–47].

Replacing $\mathrm{d}y/\mathrm{d}x$ with r, we calculate the ESS (x^*) to be

$$x^* = \frac{rb - c}{b(1 + r)}. \tag{2.3}$$

From the above equation, we can see that increasing relatedness increases the ESS value of x. Our model is agnostic to how relatedness between interacting individuals arises and therefore captures a variety of ways through which relatedness could be positive. One way to achieve higher r is through limited diffusion, because this leads to identical copies of the gene being more likely to find themselves near each other [14]. Thus, our result captures previous claims that limited diffusion would favour cooperation between replicators. Equation (2.3) is analogous to the ESS identified in Frank's [30] paired suicide model, but with an arbitrary cost of cooperation.

We found that cooperation evolves ($x^* > 0$) when $rb - c > 0$, which is the classic result known as Hamilton's [14] rule. Hamilton's rule is a relatively general result stating that a cooperative trait will evolve if the cost, c, is outweighed by the benefit, b, weighted by relatedness, r [48]. This analysis, therefore, confirms that we can think of kin selection as the reason limited diffusion favours cooperation between replicators, in exactly the same way as kin selection is usually applied to explain cooperation in other taxa, such as bacteria and animals.

3. Population structure

(a) Island model

Our above model showed that high relatedness favours cooperation, but left open the mechanism by which high relatedness is generated. One possibility is through limited diffusion of offspring copies [14], as has been argued for a wide range of organisms, including trans-replicases [12]. We test this idea by explicitly modelling population structure in an infinite island model. This is a standard approach to modelling population structure in evolutionary biology and is slightly different from a lattice model. In a lattice model we explicitly track distance, such that individuals might be further or closer apart. In an island model, we do not track distance, but instead allow individuals to stay in one place or move arbitrarily far away. The island model has been shown to give similar results to more explicit lattice and stepping stone structures [49].

Our infinite population is now subdivided into an infinite number of patches, or islands. For example, we can imagine that groups of replicators are isolated in crevices, on separate rocks, or even in droplets [4,50]. These patches have limited resources (e.g. nucleotides), such that they contain N individuals. Individual replicators interact within these patches, and these interactions determine their fitnesses or the number of offspring copies they produce.

Offspring are produced in a single generation, or round, of replication, and offspring copies diffuse to a distant patch with probability $(1 - s)$. Otherwise, with probability s, they stay on the same patch. Offspring that remain compete randomly among themselves and with new arrivees from other patches for the N available spots, and those that do not secure a spot die. Thus, an individual's fitness determines the chances that the next generation is made up of its offspring. Dispersers, similarly, compete on their new patch with other dispersers and residents for the N spots on that patch. 'Dispersal' and 'diffusion' are usually used in the kin selection and replicator literature, respectively, to mean the same thing ($(1 - s)$)—we use diffusion for replicators.

Biological models often assume that generations are discrete. This means that when offspring copies are produced, parent copies die, such that each new generation is made up exclusively of offspring. This may not be realistic for simple replicators. Thus, we allow some proportion, k, of parent individuals to survive into the next generation. Survivors maintain their spot on a patch (for example, because they are bound to one of the free binding sites). As a result, all offspring individuals are competing for the $1 - k$ fraction of free spots on a patch.

In some ways, the patches in our model are similar to cells, in that they allow associations to build up between individuals. However, they are distinct from cells in that offspring disperse independently. The diffusion rate is fixed in this model, which is justified if it is a function of an extrinsic factor (e.g. displacement by movement of the surrounding water), or a non-varying intrinsic factor (e.g. a chemical bond that is independent of mutation). k is fixed for similar reasons.

An individual's fitness now depends on whether it diffuses (with probability $(1 - s)$) or stays (s), because this will determine the individual's competitive arena. If an individual disperses, its fitness is proportional to the population average fitness, which we assume to be 1 (the population is neither growing or shrinking). If it stays, its fitness is relative to the average fitness on the patch. To determine the average fitness on the patch, it will be necessary to take into account the average phenotype of the whole patch, including the focal individual, which is equal to $(y(N - 1) + x)/N$, but, which, for simplicity we will denote Z. After diffusion, the number of individuals on a patch is equal to the number of individuals produced on a patch that stay (with probability s) plus the number of individuals arriving from elsewhere ($(1 - s)N$). So the total number of offspring on a patch after diffusion is

$$\begin{aligned} &s[N + N(bZ(1 - Z) - cZ)] + (1 - s)N \\ &\quad = N[1 + s(bZ(1 - Z) - cZ)]. \end{aligned} \tag{3.1}$$

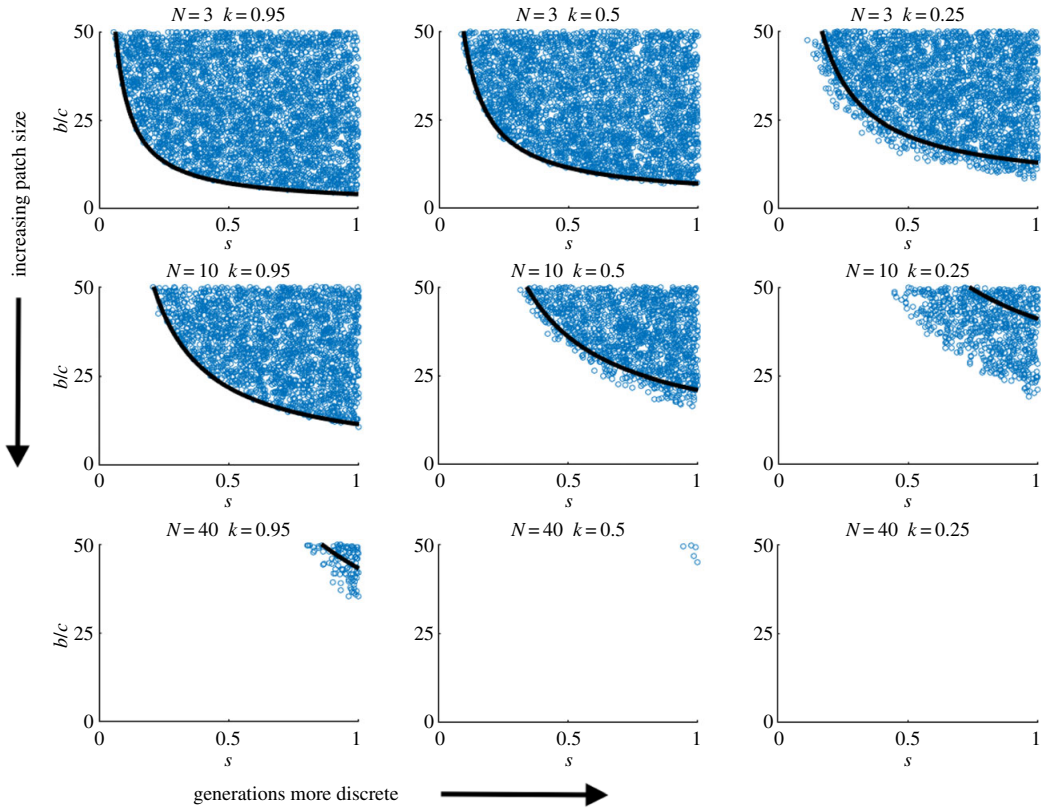

Figure 2. The condition for the evolution of cooperation when generations overlap ($k, s > 0$.) The y-axis shows the benefit to cost ratio of cooperation (b/c), and the x-axis shows the staying rate (s). The region above the solid black line is where cooperation can evolve. Open blue circles show ESS values identified from numerically solving the equilibrium determined from equation (3.2) in the main text. The benefit to cost ratio captures the degree to which a replicase can increase the replication rate of a template, relative to the associated decrease in autocatalytic replication that results from acting as an enzyme. Cooperation is more likely to be favoured by lower diffusion (increasing s); greater overlap in generations (increasing k) and smaller patch sizes (decreasing N).

These offspring then compete for the available $(1-k)$ fraction of places on the patch. This allows us to write the fitness of an individual in terms of whether it stays or disperses, as a function of x and y (remember that Z is a function of x and y):

$$w(x,y) = (1-k)(1-s)(1-cx+(1-x)by) + (1-k)s\frac{1-cx+(1-x)by}{1+s(bZ(1-Z)-cZ)} + k. \tag{3.2}$$

From this equation, we can use the Taylor–Frank approach to calculate the equilibrium strategy. However, the resulting equation is not analytically tractable. Instead, if we assume b and c are small, we can write a first-order approximation of this function that can be solved analytically. We also solved the Taylor–Frank equilibrium equation determined from equation (3.2) numerically, and found that relaxing the assumption of small b and c does not change the results (figures 2 and 3). The first-order approximation is

$$w(x,y) = (1-k)[1-cx+(1-x)by-(bZ(1-Z) - cZ)s^2] + k. \tag{3.3}$$

The fitness components in this equation have a simple biological interpretation. The terms on the left (inside the square brackets) capture the primary consequences of exhibiting the cooperative behaviour, as in the simpler model (equation (2.1)). Specifically, given an individual is cooperative, it incurs a cost, c $(-cx)$ and given it is not cooperative, it receives a benefit, b from cooperative partners $((1-x)by)$. The terms on the right capture how cooperation leads to an increase in the local competition. Specifically, extra offspring produced by the average of the trait (Z) on the patch displace the focal individual, given both the extra offspring and the focal individual remain on the patch (with probability s^2) $((bZ(1-Z)-cZ)s^2)$. This model is analogous to a haploid asexual model of others-only cooperation like that found in Taylor & Irwin [25]. As replicases cannot receive benefits, we are modelling what has been called a negatively synergistic game [51].

Using the Taylor–Frank approach, we can write the inclusive fitness effect as

$$\frac{dw}{dx} = [(1-k)(-c-by-s^2(G))] + r[(1-k)(b(1-x) - s^2(H))]. \tag{3.4}$$

The terms in the first set of square brackets are the direct effects of cooperation and the terms in the second set capture the indirect effects mediated through relatives. $G = -c/N + b(1-Z)/N - bZ/N$ and $H = (N-1)(c+b(2Z-1))/N$

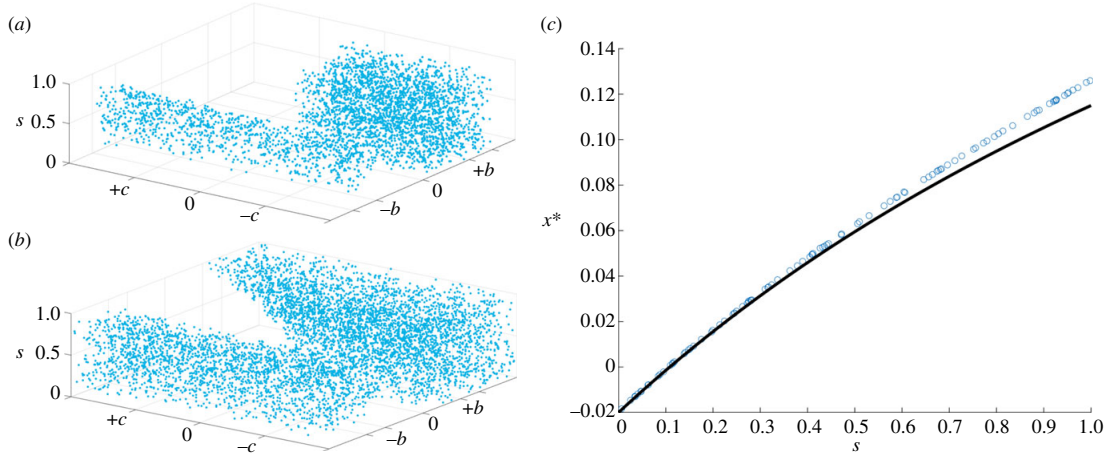

Figure 3. Possible ESS values determined numerically. The figure shows the results from numerically solving the equilibrium determined from equation (3.2) in the main text. (*a*) Discrete generations ($k = 0$, $N = 20$). Blue dots indicate candidate ESSs for different parameter values. If (*b*,*c*) are both positive, there are no possible ESS values and so cooperation cannot evolve. (*b*) The case of overlapping generations ($k = 0.9$, $N = 20$). Now, in the region where (*b*) and (*c*) are both positive, if s is high (limited diffusion), there are positive candidate ESS values, and so cooperation can evolve. (*c*) Quantitative fit between numerical (open circles) and analytical (solid line) solutions for small *b*,*c*. $k = 0.9$, $N = 5$, $b = 0.5$, $c = 0.01$.

capture the secondary effects of extra offspring that stay on the natal patch, and are decreasing functions of Z. From this, we can calculate the ESS to be

$$x^* = \frac{cN - bNr + bs^2 - cs^2 - brs^2 + crs^2 + bNrs^2 - cNrs^2}{b(-N - Nr + 2s^2 - 2rs^2 + 2Nrs^2)}. \tag{3.5}$$

This gives a solution in terms of relatedness (r) and diffusion rate ($1 - s$), but we expect r to depend on s. Limited diffusion (increasing s) should increase relatedness (r). Specifically, r is determined by the diffusion rate, the survival rate, k and the patch size, N. We can calculate r in terms of these parameter values, and plug this value for r back into equation (3.5) (see appendix A for details). This closes the model [23,26] to give the equilibrium value

$$x^* = \frac{2bks - c(2ks + N(1 + k + s - ks))}{b(1 + k)N - bs(k(N - 4) - N)}. \tag{3.6}$$

(b) Discrete generations

First, we consider the specific case of discrete generations ($k = 0$), which is the simplest possible case. When generations are discrete, we find that

$$x^* = -\frac{c}{b}. \tag{3.7}$$

This equation shows that, in the case of discrete generations, diffusion has no effect on the ESS value of cooperation—the parameter s is not in equation (3.3). Furthermore, that under limited diffusion, cooperation cannot evolve (c and b are positive, and the direction of selection at $x^* = 0$ is negative, making pure templates the stable boundary condition). This result echoes the classic result by Taylor [23,24], which showed that, while limited dispersal increases relatedness, this effect is exactly offset by the corresponding increase in local competition. This can be seen in our equation (3.4), by the two ways in which s determines the inclusive fitness effect of cooperation. Increasing s raises r, and therefore increases the indirect benefits gained by cooperating. However, increasing s also leads to the losses owing to H and G (extra offspring on the patch) being more heavily weighted. In the case of $k = 0$, these two effects exactly cancel.

(c) Overlapping generations

We now consider when there is some degree of overlapping generations ($k > 0$). In this case, the condition for cooperation to evolve becomes

$$2ksb - c(2ks + N(1 + k + s - ks)) > 0. \tag{3.8}$$

If k and s are both greater than zero—that is, if there is some degree of overlapping generations and limited diffusion—cooperation can evolve. This is because increasing k raises r, relatedness, and therefore increases the indirect benefits of cooperation without increasing the competitive effects of extra offspring (equation (3.4)). Consequently, decreasing diffusion rate ($1 - s$) and increasing survivorship (k) tend to favour cooperation (figure 2). Increasing the benefit of cooperation, b, and decreasing the cost of cooperation, c, make it easier for cooperation to evolve.

Decreasing patch size (N) makes it easier for cooperation to evolve. This is because the larger the patch size, the lower the average relatedness on a patch (equation (A 1)). One caveat is that we assume, deterministically, that each patch contains both templates and replicases. As N gets smaller, stochastic variation in the patch composition make this less likely to hold. In the extreme, if $N = 1$, then the patch could only contain a template or a replicase, but not both. Consequently, replicases would be paying the cost of cooperation, when there are no templates to gain the benefit. This stochastic effect would be reduced or removed if cooperation is conditional upon being in a patch where there are templates. An analogous problem of stochasticity in small patch sizes has been considered with sex allocation in structured populations (local mate competition) [45].

In our above model, we assumed that survivors maintain their spot on a patch. This is reasonable if, for example, once a replicator finds a binding site it remains there until death. Alternatively, we might allow survivors to remain on the patch but to compete equally with offspring for a place. This would be reasonable if, for example, offspring can 'bump' adults from a patch. A third possibility is that survivors can disperse along with offspring, and compete globally—this might occur if during each replication event replicators are dislodged from their binding site. We show in appendix A that neither allowing for survivors to compete for sites nor allowing survivors to disperse qualitatively alters the results, although both changes make cooperation more difficult to evolve.

4. Discussion

We have used the analytical tools of social evolution theory to model a simple replicating molecule scenario: transacting replicases. We have shown that cooperation between replicators can be understood as evolving via the process of kin selection through limited diffusion. However, we have also shown that limited diffusion on its own does not favour cooperation (equation (3.7)). Instead, an additional life-history detail of simple replicators is needed—that of overlapping generations (figure 2).

Our social evolution model illustrates two points about replicators. First, we can view limited diffusion as favouring cooperation in simple replicators via kin selection. Consequently, the factor favouring cooperation in trans-replicases: (i) links to a large existing theoretical literature [14,21–23], and (ii) is the same factor that has been previously shown to favour cooperation in a range of organisms including birds, mammals, insects and microbes [15–20,52,53]. By clarifying these links across taxa, we can simplify our understanding of life, rather than having to provide different explanations for different cases. We are not saying cooperation in replicators has to be conceptualized via kin selection, just that it can be useful to do so.

Second, both limited diffusion and overlapping generations are required to favour cooperation. Limited diffusion leads to a build-up of relatedness, which favours cooperation [14]. But at the same time, limited diffusion leads to increased competition between relatives, which disfavours cooperation [23,31,54]. Overall, in the simplest possible scenario, these two effects exactly cancel (equation (3.7)). However, we found that the addition of overlapping generations allows limited diffusion to favour cooperation (figure 2). When generations overlap, this increases relatedness within patches, but without increasing competition between relatives, because offspring still diffuse to the same extent [25]. Specifically, increasing overlap (k) raises relatedness ((r), and therefore increases the indirect benefits of cooperation without increasing the competitive effects of extra offspring (equation (3.4)). Consequently, when there is both limited diffusion and overlapping generations, the build-up of relatedness outweighs the increased competition between relatives, such that cooperation is favoured (figure 2).

There are many ways to model social behaviours. One decision is whether to assume discrete strategies, such as 'cooperators' and 'cheats', or to allow for continuous strategies, ranging, for example, from completely selfish to completely cooperative [29,39,55]. The assumption of continuous strategies is clearly valid for animals, where traits are determined by multiple genes, but simple replicators might only have a limited number of strategies open to them by mutation. Another decision is whether to develop explicit simulations or analytical models. Simulations allow greater detail to be incorporated, which can be especially useful when considering specific systems or species. By contrast, the analytical approach usually used in kin selection models tends to be more streamlined, sacrificing details and precision for clarity and generality [29,31]. Further, the kin selection approach offers a heuristic which non-mathematicians can apply across a range of organisms [45,47]. Rather pleasingly, in the replicator scenario examined here, the different approaches make the same qualitative prediction [12].

The route from independent replicators to the first genomes probably involved two kinds of cooperation. Early cooperation could have been between genetic relatives, or replicators of the same type. However, early genomes were probably too simple to copy themselves accurately, and yet inaccurate replication prevented genomes from getting large enough to improve their accuracy [1]. In order for the genome to overcome this 'error threshold in replication', it is probable that different types of replicators needed to cooperatively copy each other. Individual replicators could remain small enough to be copied accurately, but the collection of replicators could become large enough to produce the kind of enzyme machinery needed for accuracy [2]. We have modelled the first kind of cooperation—between replicators of the same 'type'—and have shown that this can be understood as evolving via kin selection [28]. Cooperation between different types, however, would have required some factor to align the interests of unrelated replicators.

To conclude, although we have phrased our model in terms of a specific replicator system, the trans-replicase, we expect our predictions to hold more generally for other types of replicators. We do not yet know the actual biology of the earliest life forms. But, while many higher organisms may adopt a system of discrete generations, we would expect overlapping generations to be a feature of all simple replicators. Our results, then, would apply to various possible routes through which simple replicators could come together to cooperate.

Data accessibility. This article has no additional data.

Author's contributions. S.R.L. and S.A.W. contributed to conception, modelling and write up of the manuscript. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. S.R.L. is funded by The Clarendon Fund, Hertford College and the Natural Environment Research Council.

Acknowledgements. We thank Miguel dos Santos, Guy Cooper, Matishalin Patel, Tom Scott, Asher Leeks, Paul Higgs, Peter Taylor, Geoff Wild and one anonymous reviewer for very helpful discussion and/or comments; and Magdalen College for emergency housing. This paper was inspired by Shay *et al.* [12].

Appendix A

(a) Writing relatedness in terms of model parameters

We start by determining the relatedness, at equilibrium, of a focal individual to a random member of its patch, drawn with replacement. This is known as whole-group relatedness (denoted by R), because it includes the focal individual, in contrast with others-only relatedness, which does not include

the focal individual [42]. Note that in our model, we are dealing with r, others-only relatedness, because y is the average of the individuals on the patch, excluding the focal individual. We can write R (whole-group relatedness), the relatedness between two individuals drawn randomly from a patch with replacement, as the probability that those two individuals are the same individual $(1/N)$, and thus have relatedness 1, plus the probability that those two individuals are not the same $((N-1)/N)$, and thus have the relatedness of two random individuals drawn without replacement, or others-only relatedness, r:

$$R = \frac{1}{N} + \frac{N-1}{N} r. \tag{A 1}$$

Now we take two individuals (without replacement) on the same patch with relatedness r, and determine the relatedness of their representatives in the previous generation. With chance k^2 they are both survivors from the previous generation, in which case their relatedness is the same (r). With chance $2k(1-k)$ one is a survivor and the other is a new offspring, which is native with probability s, in which case their relatedness is R. Else, with chance $(1-k)^2$ they are both new offspring, are both native with probability s^2, and thus have relatedness R. We can write others-only relatedness between two individuals in the current generation as equal to

$$r_t = k^2 r_{t-1} + 2k(1-k)sR_{t-1} + (1-k)^2 s^2 R_{t-1}. \tag{A 2}$$

Here r_t is relatedness in the current generation, or time step, and r_{t-1} and R_{t-1} are others-only and whole-group relatednesses, respectively, in the previous one. Setting $r_t = r_{t-1}$ we find the equilibrium others-only relatedness. Plugging into equation (A 1), we find the equilibrium value of whole-group relatedness, R^*, to be

$$R* = \frac{1+k}{n+kn+2ks-2kns+s^2-ks^2-ns^2+kns^2}. \tag{A 3}$$

This equation for relatedness was identified by Taylor & Irwin [25]. However, here we are modelling an others only-trait, and thus require others-only relatedness, r. RN gives us the number of relatives on our patch. Subtracting the focal individual, and dividing by the total number of remaining individuals $(N-1)$, gives us r^*:

$$r* = \frac{n(1+k)/((1+k)n-(n-1)(2k+s-ks)s)-1}{n-1}. \tag{A 4}$$

Plugging into equation (3.5) gives us equation (3.6).

(b) Allowing survivors to remain and compete for patch sites or disperse globally

Our original model assumed that surviving parents maintained their places on a patch, meaning offspring competed for the remaining $1-k$ fraction of available sites. Here we relax this assumption. First, we allow survivors to remain on their patch, but compete equally with offspring for available sites. Using the relatedness recursion in equation (A 2), we calculate the ESS (assuming small b and c) to be

$$\begin{aligned} x* = {} & (c(-1+k)(1+k)^2 n - k(1+k)((b-c)(-1+2k)+3c(-1+k)n)s \\ & +(-1+k)((b-c)k(1+3k)+c(-1+3k^2)n)s^2-(-1+k)^2 \\ & k(b+c(-1+n))s^3) \Big/ (b((-1+k)n(-1+k(-1+s)) \\ & (-1+k(-1+s)-s)(-1+s)-2ks(-1+k(1+2k)+s+(2-3k) \\ & ks+(-1+k)^2 s^2))). \end{aligned} \tag{A 5}$$

Next, we allow survivors to disperse along with offspring. This requires a new relatedness recursion, which we write as

$$r_t = k^2 s^2 r_{t-1} + 2k(1-k)s^2 R_{t-1} + (1-k)^2 s^2 R_{t-1}. \tag{A 6}$$

We now determine the ESS to be

$$\begin{aligned} & (-c(-1+k)n - c(-1+k)kns + ((-b+c)k(-1+(-1+k)k) \\ & +c(-1+k^2)n)s^2 + (-1+k)k(b-bk^2+c(-1+k^2+n))s^3 \\ & +k((b-c)(1+(-2+k)k(1+k))-c(-1+k)n)s^4) \\ & \Big/ (b((-1+k)n(-1+k(-1+s)s)(-1+s^2) \\ & +2ks^2(1+k-k^2-(-1+k)^2(1+k)s+(1+(-2+k)k(1+k))s^2))). \end{aligned} \tag{A 7}$$

If $k = 0$ (discrete generations), both equations (A 5) and (A 7) revert to equation (3.7) in the text, and cooperation cannot evolve. However, given some degree of overlapping generations and limited diffusion, cooperation can evolve, although the condition is now more stringent.

References

1. Eigen M. 1971 Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465–523. (doi:10.1007/BF00623322)
2. Eigen M, Schuster P. 1977 A principle of natural self-organization. *Naturwissenschaften* **64**, 541–565. (doi:10.1007/BF00450633)
3. Smith JM, Szathmary E. 1995 *The major transitions in evolution*. Oxford, UK: Oxford University Press.
4. Higgs PG, Lehman N. 2015 The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17. (doi:10.1038/nrg3841)
5. Boerlijst MC, Hogeweg P. 1991 Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Phys. D Nonlinear Phenom.* **48**, 17–28. (doi:10.1016/0167-2789(91)90049-F)
6. Boerlijst MC, Hogeweg P. 1995 Spatial gradients enhance persistence of hypercycles. *Phys. D Nonlinear Phenom.* **88**, 29–39. (doi:10.1016/0167-2789(95)00178-7)
7. Cronhjort MB, Blomberg C. 1997 Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Phys. D Nonlinear Phenom.* **101**, 289–298. (doi:10.1016/S0167-2789(97)87469-6)
8. McCaskill JS, Füchslin RM, Altmeyer S. 2001 The stochastic evolution of catalysts in spatially resolved molecular systems. *Biol. Chem.* **382**, 1343–1363. (doi:10.1515/BC.2001.167)
9. Szabó P, Scheuring I, Czárán T, Szathmáry E. 2002 *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**, 340–343. (doi:10.1038/nature01187)
10. Sardanyés J, Solé RV. 2007 Spatio-temporal dynamics in simple asymmetric hypercycles under

weak parasitic coupling. *Phys. D Nonlinear Phenom.* **231**, 116–129. (doi:10.1016/j.physd.2007.04.009)

11. Bianconi G, Zhao K, Chen IA, Nowak MA. 2013 Selection for replicases in protocells. *PLoS Comput. Biol.* **9**, e1003051. (doi:10.1371/journal.pcbi.1003051)
12. Shay JA, Huynh C, Higgs PG. 2015 The origin and spread of a cooperative replicase in a prebiotic chemical system. *J. Theor. Biol.* **364**, 249–259. (doi:10.1016/j.jtbi.2014.09.019)
13. Wächtershäuser G. 1988 Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev.* **52**, 452–484.
14. Hamilton WD. 1964 The genetical theory of social behavior. I and II. *J. Theor. Biol.* **7**, 1–52. (doi:10.1016/0022-5193(64)90038-4)
15. Griffin AS, West SA, Buckling A. 2004 Cooperation and competition in pathogenic bacteria. *Nature* **430**, 1024–1027. (doi:10.1038/nature02744)
16. Hughes WOH, Oldroyd BP, Beekman M, Ratnieks FLW. 2008 Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–1216. (doi:10.1126/science.1156108)
17. Diggle SP, Gardner A, West SA, Griffin AS. 2007 Evolutionary theory of bacterial quorum sensing: when is a signal not a signal? *Phil. Trans. R. Soc. B* **362**, 1241–1249. (doi:10.1098/rstb.2007.2049)
18. Kuzdzal-Fick JJ, Fox SA, Strassmann JE, Queller DC. 2011 High relatedness is necessary and sufficient to maintain multicellularity in dictyostelium. *Science* **334**, 1548–1551. (doi:10.1126/science.1213272)
19. Lukas D, Clutton-Brock T. 2012 Cooperative breeding and monogamy in mammalian societies. *Proc. R. Soc. B* **279**, 2151–2156. (doi:10.1098/rspb.2011.2468)
20. Fisher RM, Cornwallis CK, West SA. 2013 Group formation, relatedness, and the evolution of multicellularity. *Curr. Biol.* **23**, 1120–1125. (doi:10.1016/j.cub.2013.05.004)
21. West SA, Pen I, Griffin AS. 2002 Cooperation and competition between relatives. *Science* **296**, 72–75. (doi:10.1126/science.1065507)
22. Lehmann L, Rousset F. 2010 How life history and demography promote or inhibit the evolution of helping behaviours. *Phil. Trans. R. Soc. B* **365**, 2599–2617. (doi:10.1098/rstb.2010.0138)
23. Taylor PD. 1992 Altruism in viscous populations—an inclusive fitness model. *Evol. Ecol.* **6**, 352–356. (doi:10.1007/BF02270971)
24. Taylor PD. 1992 Inclusive fitness in a homogeneous environment. *Proc. R. Soc. Lond. B* **249**, 299–302. (doi:10.1098/rspb.1992.0118)
25. Taylor PD, Irwin AJ. 2000 Overlapping generations can promote altruistic behavior. *Evolution* **54**, 1135–1141. (doi:10.1111/j.0014-3820.2000.tb00549.x)
26. Gardner A, West SA. 2006 Demography, altruism, and the benefits of budding. *J. Evol. Biol.* **19**, 1707–1716. (doi:10.1111/j.1420-9101.2006.01104.x)
27. Lehmann L, Perrin N, Rousset F, Day T. 2006 Population demography and the evolution of helping behaviors. *Evolution* **60**, 1137–1151. (doi:10.1111/j.0014-3820.2006.tb01193.x)
28. Frank SA. 1994 Kin selection and virulence in the evolution of protocells and parasites. *Proc. R. Soc. Lond. B* **258**, 153–161. (doi:10.1098/rspb.1994.0156)
29. Taylor PD, Frank SA. 1996 How to make a kin selection model. *J. Theor. Biol.* **180**, 27–37. (doi:10.1006/jtbi.1996.0075)
30. Frank SA. 1997 Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *J. Theor. Biol.* **189**, 307–316. (doi:10.1006/jtbi.1997.0516)
31. Frank SA. 1998 *Foundations of social evolution*. Princeton, NJ: Princeton University Press.
32. Inoue T, Orgel LE. 1983 A nonenzymatic RNA polymerase model. *Science* **219**, 859–862. (doi:10.1126/science.6186026)
33. Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP. 2001 RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325. (doi:10.1126/science.1060786)
34. Zaher HS, Unrau PJ. 2007 Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026. (doi:10.1261/rna.548807)
35. Rajamani S, Ichida JK, Antal T, Treco DA, Leu K, Nowak MA, Szostak JW, Chen IA. 2010 Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.* **132**, 5880–5885. (doi:10.1021/ja100780p)
36. Wochner A, Attwater J, Coulson A, Holliger P. 2011 Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212. (doi:10.1126/science.1200752)
37. Attwater J, Wochner A, Holliger P. 2013 In-ice evolution of RNA polymerase ribozyme activity. *Nat. Chem.* **5**, 1011–1018. (doi:10.1038/nchem.1781)
38. Fisher Ronald A. 1930 *The genetical theory of natural selection*. Oxford, UK: University Press Google Scholar.
39. MaynardSmith J, Price GR. 1973 The logic of animal conflict. *Nature* **246**, 15–18. (doi:10.1038/246015a0)
40. Hamilton WD. 1970 Selfish and spiteful behaviour in an evolutionary model. *Nature* **228**, 1218–1220. (doi:10.1038/2281218a0)
41. Grafen A. 1985 A geometric view of relatedness. *Oxf. Surv. Evol. Biol.* **2**, 28–89.
42. Pepper JW. 2000 Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**, 355–368. (doi:10.1006/jtbi.2000.2132)
43. Sachs JL, Mueller UG, Wilcox TP, Bull JJ. 2004 The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160. (doi:10.1086/383541)
44. West SA, Griffin AS, Gardner A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672. (doi:10.1016/j.cub.2007.06.004)
45. West S. 2009 *Sex allocation*. Princeton, NJ: Princeton University Press.
46. Abbot P *et al.* 2011 Inclusive fitness theory and eusociality. *Nature* **471**, E1–E4. (doi:10.1038/nature09831)
47. Bourke AFG. 2014 Hamilton's rule and the causes of social evolution. *Phil. Trans. R. Soc. B* **369**, 20130362. (doi:10.1098/rstb.2013.0362)
48. Gardner A, West SA, Wild G. 2011 The genetical theory of kin selection. *J. Evol. Biol.* **24**, 1020–1043. (doi:10.1111/j.1420-9101.2011.02236.x)
49. Comins HN, Hamilton WD, May RM. 1980 Evolutionarily stable dispersal strategies. *J. Theor. Biol.* **82**, 205–230. (doi:10.1016/0022-5193(80)90099-5)
50. Zwicker D, Seyboldt R, Weber CA, Hyman AA, Jülicher F. 2016 Growth and division of active droplets provides a model for protocells. *Nat. Phys.* **13**, 408–413. (doi:10.1038/nphys3984)
51. Queller DC. 1985 Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**, 366–367. (doi:10.1038/318366a0)
52. Cornwallis CK, West SA, Griffin AS. 2009 Routes to indirect fitness in cooperatively breeding vertebrates: kin discrimination and limited dispersal. *J. Evol. Biol.* **22**, 2445–2457. (doi:10.1111/j.1420-9101.2009.01853.x)
53. Cornwallis CK, West SA, Davis KE, Griffin AS. 2010 Promiscuity and the evolutionary transition to complex societies. *Nature* **466**, 969–972. (doi:10.1038/nature09335)
54. Queller DC. 1994 Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70–73. (doi:10.1007/BF01237667)
55. Grafen A. 1979 The hawk-dove game played between relatives. *Anim. Behav.* **27**, 905–907. (doi:10.1016/0003-3472(79)90028-9)

3

Kin Selection in the RNA World

life

MDPI

Article

Kin Selection in the RNA World

Samuel R. Levin * and Stuart A. West

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK; stuart.west@zoo.ox.ac.uk

* Correspondence: samuel.levin@zoo.ox.ac.uk

Received: 15 October 2017; Accepted: 30 November 2017; Published: 5 December 2017

Abstract: Various steps in the RNA world required cooperation. Why did life's first inhabitants, from polymerases to synthetases, cooperate? We develop kin selection models of the RNA world to answer these questions. We develop a very simple model of RNA cooperation and then elaborate it to model three relevant issues in RNA biology: (1) whether cooperative RNAs receive the benefits of cooperation; (2) the scale of competition in RNA populations; and (3) explicit replicator diffusion and survival. We show: (1) that RNAs are likely to express partial cooperation; (2) that RNAs will need mechanisms for overcoming local competition; and (3) in a specific example of RNA cooperation, persistence after replication and offspring diffusion allow for cooperation to overcome competition. More generally, we show how kin selection can unify previously disparate answers to the question of RNA world cooperation.

Keywords: RNA cooperation; kin selection; RNA world; Hamilton's rule; limited diffusion; origin of the genome; scale of competition; modelling the origin of life

1. Introduction

Life very likely began as simple replicating RNA molecules [1–3]. These first replicators were capable of little more than making copies of themselves. However, the last universal common ancestor already contained a complex genome, wrapped inside a cell, capable of varied metabolic and replicative tasks. Replicators in the RNA world then had many obstacles to overcome. Molecules had to successfully copy themselves and each other. Different kinds of ribozymes, such as polymerases and synthetases, had to evolve and stably persist. Independent replicators had to come together to form the first genomes. Each of these steps involved various biochemical hurdles, and most of these biochemical puzzles remain largely unsolved.

However, in addition to posing biochemical problems, many of the key steps in the evolution of the RNA world are also problems of cooperation [4–7]. For example, in various plausible RNA world scenarios, molecules act as enzymes to increase the replication rate of other molecules (Figure 1). This can be considered a cooperative trait, because by acting as enzymes, these molecules reduce their own replication rate to help copy others. A selfish mutant that receives the benefits of the enzymatic activity of others, but does not act as an enzyme itself, would have higher fitness. Consequently, all else being equal, we expect selfish molecules to outcompete cooperative ones. The problem is simple: Why would a replicating molecule help copy others instead of selfishly copying itself as fast as it could? Similar problems arise in synthestases cooperatively producing nucleotides and independent replicators coming together to form the first genomes. Each of these issues is a different problem of cooperation in the RNA world, and a number of explanations have been put forward to resolve these problems, including limited diffusion, primitive cells, and spiral waves [5,6,8–17].

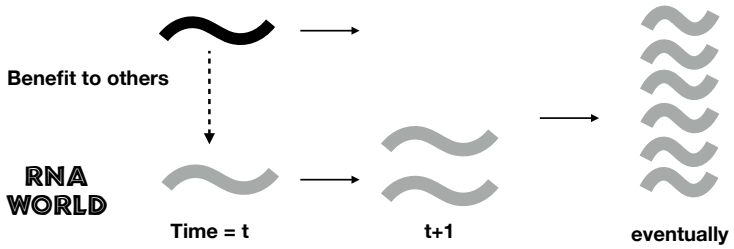

Figure 1. The problem of cooperation in the RNA world. A cooperator (black squiggle) provides a benefit to other individuals (grey squiggle), increasing their relative replicative success at a cost to their own relative success. Over time, all else being equal, individuals that do not incur this cost but receive the benefits have higher replicative success, or fitness, and become better represented in the population. How, then, does cooperation evolve?

The links between these different suggested solutions for the problem of cooperation in the RNA world are not clear. Are they different explanations, or can we, instead, identify an overarching framework that links them all? It is useful here to make a comparison of the literature on the evolution of cooperation in higher organisms, ranging from animals to bacteria. Over the last 50+ years, work in this area of 'social evolution' has produced relatively unified theoretical and empirical literature that can explain cooperation across the tree of life [18–20]. One theory that could be especially relevant to the RNA world is kin selection [6,9,17]. Hamilton showed that cooperation can be explained if it is directed towards relatives [21]. Natural selection favours genes that are better able to get copies of themselves into the next generation. Hamilton's kin selection theory highlights that a gene can get a copy of itself into the next generation by either increasing the replication of the individual it is in, or by increasing the replication of other individuals that carry copies of that gene.

Kin selection requires that cooperation be directed towards relatives. While this can involve mechanisms to discriminate kin from non-kin, it can also work via limited dispersal keeping relatives together. For example, when bacteria grow clonally, a cell will be tend to be surrounded by genetically identical cells, facilitating cooperation. This could potentially be important in the RNA world if limited diffusion keeps copies of the same molecule together (Figure 2). Indeed, several of the models developed to explain cooperation in the RNA world appear similar to previously developed kin selection models. If cooperation in the RNA world can be explained by kin selection, then this would simplify our picture of the world, unifying existing RNA models and showing how the same process can drive biochemical and organism level evolution.

Our aim is to test the utility of using existing kin selection methodologies to explain cooperation in the RNA world. We use a game theory approach to determine under what conditions cooperation would be evolutionarily stable, and hence be favoured by natural selection. This approach is deliberately simple, abstracting away biological details, to focus on key parameters that are likely to be important across different systems. We start with the simplest possible model, and then elaborate by adding in details that could have been important in the RNA world. Our more specific questions are: (1) Why would cooperation have been favoured in the RNA world? (2) How will different RNA biochemistries influence the evolution of cooperation? (3) Can we gain anything from applying the kin selection approach to the RNA world?

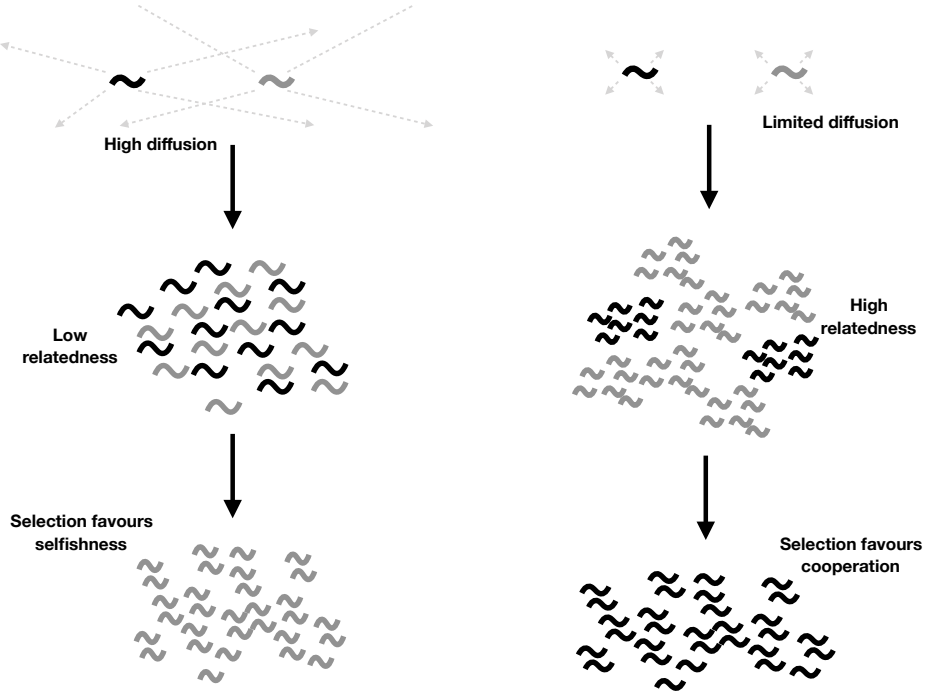

Figure 2. Limited diffusion in the RNA world. Cooperative RNAs are depicted as black squiggles, and selfish ones as grey squiggles. High diffusion leads to a well mixed population (low relatedness), which favours the evolution of selfishness. Limited diffusion leads to high relatedness. Cooperators are more likely to encounter other cooperators, and selfish individuals are unlikely to encounter cooperative ones to exploit. Selection favours cooperation.

2. Results

2.1. A Simple Model of RNA Cooperation

We start by developing a simple model of cooperation in RNA molecules using standard kin selection techniques [22]. We deliberately avoid tying the model to a specific RNA system, keeping it general to capture a broad range of possible systems. A similar approach has recently been taken with viruses [23]. We imagine that RNA molecules have some potentially cooperative trait that benefits the other RNA molecules with whom they are interacting (social partners). For example, an RNA replicator might act as a cooperative enzyme, increasing the replication rate of the local replicators. However, at the same time, this cooperation comes at a cost to the individual performing the cooperation, reducing their replication rate. For example, a cost could arise because acting as an enzyme reduces the amount of time a molecule is available in template form for other molecules [16,24]. Consequently, there is a trade-off here, with cooperation benefiting the group, but being costly to the individual.

In this case, the replication rate, or fitness, of an RNA molecule will be a function of its own level of cooperation (the individual cost) and the local level of cooperation amongst the RNA molecules it is interacting with (the group benefit). We assume that the focal RNA molecule has phenotype x and that the average phenotype of the social partners it is interacting with is y. The phenotype could represent the likelihood (bounded between zero and 1) or amount (unbounded) of cooperation. For example, x could be the probability, between 0 and 1, of becoming a cooperative enzyme. A simple way to model the costs and benefits of cooperation is to assume that our focal RNA molecule has a baseline replication rate of W_b, which will be reduced by some function depending upon its own level of cooperation (C), and increased by some function of the level of cooperation in the local group of RNA molecules (B). The replication rate, w, can then be expressed as:

$$w(x, y) = W_b - C(x) + B(y). \quad (1)$$

We can think of C as the cost of the cooperative trait and B as the benefit of the trait. Equation (1) is analogous to other models that have looked at cooperation both generally and in specific systems such as microbes [22,25–29]. We now ask what strategy would be favoured by natural selection. More formally, we seek the evolutionarily stable strategy (ESS) [30]. An ESS is a strategy (x^*) for which, if all individuals in the population express it, no rare mutant variant will have a higher replication rate (Figure 3). The ESS approach focuses on phenotypes, and assumes that, within a range set by the modeller, all phenotypes are possible. For example, the likelihood that an RNA molecule cooperates could be between 0 and 1. Another way of thinking about this is that the ESS approach looks at which direction evolution would proceed by looking for the unbeatable strategy.

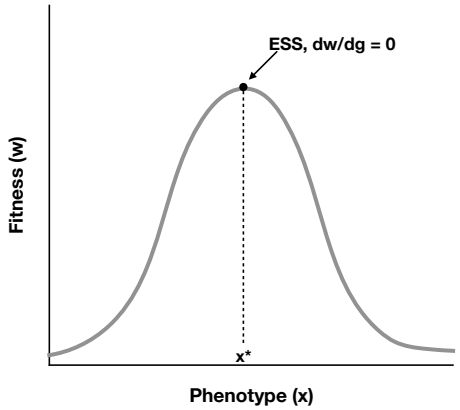

Figure 3. Visual representation of the evolutionarily stable strategy (ESS) approach. Taylor and Frank (1996) developed an approach for identifying ESSs. An equation for fitness (y-axis) as a function of phenotype (x-axis) is either derived or assumed. Natural selection will move populations towards fitness peaks. At a fitness optimum, small phenotypic variations in either direction will have lower fitness, and therefore the population will remain the same. The ESS is the phenotype (x^*) where this occurs, and for a continuously differential fitness function, this happens at $dw/dx = 0$. We expect organisms to express ESSs as a result of natural selection over time.

Candidate ESSs occur where fitness is maximized. Taylor and Frank (1996) showed that, assuming weak selection, this happens when the derivative of fitness with respect to genotype is zero [31]:

$$\frac{dw}{dg} = \frac{\partial w}{\partial x} + R\frac{\partial w}{\partial y} = 0. \quad (2)$$

R is genetic relatedness, which is a measure of genetic similarity [21,32,33]. What does relatedness mean in the context of RNA replicators? In general, relatedness captures the likelihood that two individuals share genes, and therefore it measures an individual's vested interest (in an evolutionary sense) in others. Genetic similarity between partners can come about a number of ways. Thus, R is a very general parameter which captures all processes that generate genotypic associations between individuals.

In the case of RNAs, this association could come about, for example, through limited diffusion of offspring copies. Limited diffusion leads to identical copies finding themselves near each other, meaning R is high. Consequently, a focal molecule's genetic sequence can become better represented in the population either by copying itself or its neighbours, which are likely to be identical. In the RNA world, relatedness usually measured the proportion of interactants that had an identical sequence, but relatedness was able to arise from any heritable correlation in RNA traits. As an example, in the simplest case of a social group containing τ individuals with identical phenotypes and π with different phenotypes, $R = \frac{\tau}{\tau + \pi}$ (for τ individuals). This simple measure of relatedness allows us to capture many different possible configurations of genotypes and phenotypes in a single parameter, vastly simplifying our analysis. In this

model R is equal to $\frac{dy}{dx}$, and relatedness is a measure of phenotypic correlation—the more closely related you are to individuals, the more similar your phenotypes will be.

Solving Equation (2) for $x = y = x^*$ (a monomorphic population) gives the value of the ESS, x^*. The condition for cooperation to evolve is $x^* > 0$ (because $x = 0$ would be no cooperation). For Equation (1), assuming baseline fitness is 1, cooperation evolves when

$$RB' - C' > 0, \tag{3}$$

where $B' = \frac{\partial w}{\partial y}$ is the benefit of cooperation, $C' = \frac{\partial w}{\partial x}$ is the cost of cooperating, and R is relatedness.

Equation (3) tells us that RNA cooperation will be favoured if the marginal cost of cooperation is smaller than the marginal benefit of cooperation, weighted by relatedness between social partners. Thus, Equation (3) captures previous results that cooperation can be favoured by limited diffusion, spiral waves, or primitive cells. Limited diffusion and spiral waves lead to identical copies finding themselves near each other, which generates high R [7,17]. Primitive cells generate high R by keeping identical individuals together from one generation to the next [9].

Equation (3) also illustrates how we can think about cooperation in RNA molecules as being favoured by kin selection. If there is a high likelihood that molecules will interact with identical molecules (high R), then cooperation will more readily evolve. Specifically, Equation (3) is a classic result known as Hamilton's (1964) rule, which has been used to explain cooperation across the tree of life [21,28,32]. In particular, a role for limited dispersal in generating a positive relatedness, and hence favouring cooperation by kin selection, has been demonstrated in a range of organisms, including bacteria, slime moulds, insects, birds and mammals [34–42].

2.2. Different Types of RNA Cooperation

In the above section, we were deliberately vague about the functions B and C, to keep them general. We did this so that the model would capture a wide range of potentially cooperative traits in the RNA world. However, different types of RNA molecules will engage in different types of cooperation, and it can be useful to consider these more explicitly. One issue is that sometimes cooperators can also be recipients of the benefits of cooperation and other times they are not. For example, a cooperative polymerase might replicate a nucleotide synthetase, leading to more nucleotides in the environment, which will benefit the focal polymerase as well as all those around it. Alternatively, by acting as a replicase to increase the replication rate of others, a replicator might no longer be able to be copied itself, such that it receives none of the benefits of cooperation. Do these different forms of cooperation lead to different evolutionarily stable strategies? Does cooperation evolve to the same degree regardless of the form it takes?

We can answer these questions by extending the previous model to explicitly model the nature of the cooperative trait. We again imagine an RNA replicator can express some cooperative traits, like acting as an enzyme to increase the replication rate of others, at a cost to itself. We use the term 'cooperative' in line with other work in the field, but since in this case replicators incur a lifetime fitness cost to provide a benefit to others, this is 'altruism' in the strict sense [43]. An individual's strategy relates to the proportion of its offspring copies that express the cooperative phenotype. For example, alternate folding patterns would allow different copies to express different phenotypes. A more cooperative strategy would be one where a higher proportion of offspring acts as cooperative enzymes.

Baseline fitness is assumed to be 1. We take $0 \leq x \leq 1$ to be the proportion of a focal individual's offspring copies that cooperates. For example, if $x = 1$, all the copies of an offspring act as cooperative enzymes (complete cooperation). If $x = 0$, none of its offspring cooperate (complete selfishness). Values between 0 and 1 are considered partial cooperation. b and $0 \leq c \leq 1$ are the benefits and costs, respectively, of cooperating (where here, for simplicity, we have substituted concrete effects on replication rate for the functions in Equation (1)). For example, if an individual becomes an enzyme

that helps replicate others, it loses c from its baseline replication rate, but increases the replication rate of others around it by a factor b.

To distinguish between different forms of cooperation, we incorporate a new parameter, $0 \leq \beta \leq 1$. β measures the degree to which, given a molecule is cooperative, it can still receive the benefits of cooperation (by). For example, if folding to act as an enzyme prevented a cooperator from being replicated by other enzymes, this would be represented by $\beta \simeq 0$. Or if the cooperator mined some public good, like nucleotides, which benefits the group, this would lead to $\beta \simeq 1$, as the cooperator can receive the benefits of other molecules mining nucleotides (Figure 4).

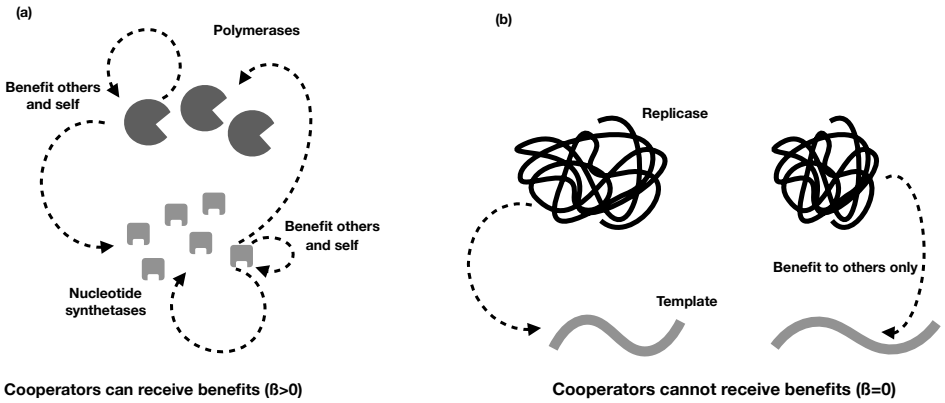

Figure 4. An illustration of the different types of cooperation in the RNA world. (**a**) Nucleotide synthetases (light grey cuboids) make nucleotides which benefit themselves, other nucleotides, and polymerases (dark grey spheroids). β is relatively high, because being a cooperator does not limit a synthetase's ability to benefit from cooperation. Similarly, polymerases can copy nucleotides and other polymerases, and they receive benefits both by making more synthetases (which leads to more nucleotides) and by being copied by other polymerases. (**b**) A cooperative replicase can copy a template, but cannot be copied by other replicases. Thus, $\beta = 0$, because being a cooperator prevents one from receiving any benefits from cooperation.

An individual's fitness will be the sum of the replication rates of the fraction x of its offspring which act as cooperators and the fraction $(1 - x)$ that is selfish:

$$w(x, y) = x(1 - c + \beta by) + (1 - x)(1 + by). \tag{4}$$

Offspring copies that do not act as cooperators have, on average, a $1 + by$ relative replication rate. Offspring that act as cooperators have a $1 - c + \beta by$ replication rate, where β measures the degree to which cooperators receive benefits. The Taylor–Frank method shows that the direction of selection is given by

$$\frac{dw}{dg} = 1 - c + \beta bx - 1 - bx + R(\beta bx + (1 - x)b). \tag{5}$$

First, we consider the extreme case where being a cooperator has no effect on an RNA's ability to receive benefits, or $\beta = 1$. In game theoretic terms this is equivalent to an additive game. This fits the scenario, for example, in which a molecule can 'mine' nucleotides from the environment, which benefits all individuals in the group, including the cooperator. In this case, Equation (5) reverts to Equation (3), and cooperation will evolve when

$$Rb - c > 0, \tag{6}$$

which again represents a simplified form of Hamilton's rule. Otherwise, if $\beta > 0$, candidate ESSs are given by

$$x^* = \frac{Rb - c}{b(1+R) - \beta b(1+R)}. \quad (7)$$

Here and for all subsequent analyses we assume that x^* is bounded between 0 and 1 (checking that the boundaries are stable when this is not true). Equation (7) is a general result for a negatively synergistic game played by RNA molecules in which being a cooperator reduces a replicator's ability to be a recipient of cooperation. We find that:

1. Regardless of whether cooperative RNAs receive benefits (the value of β), the condition for cooperation to evolve ($x^* > 0$) is $Rb - c > 0$. This tells us that the degree to which cooperators receive benefits has no effect on *whether* cooperation will evolve, although it impacts on the degree of cooperation. Regardless, higher benefits and relatedness and lower costs are favourable for RNA cooperation, confirming the more general model in the previous section.
2. In the extreme case in which cooperative RNAs receive no benefits ($\beta = 0$), the ESS value of cooperation is capped at 0.5, and only partial cooperation can evolve, because complete cooperation would mean that there were no individuals available to receive benefits. This applies, for example, in some RNA trans-replicases, in which becoming a replicase enzyme prevents individuals from being replicated by other replicases [16,44]. This result is analogous to the result obtained in Frank's (1997) model of paired sibling suicide in animals [45]. This suggests that the evolution of the cooperative enzyme that cannot receive benefits (as in [16,24,44]) is analogous to the evolution of sterility in higher organisms.

More generally, β determines the degree to which complete cooperation can evolve in RNAs (Figure 5), and provides a parameter that applies across forms of cooperation in the RNA world. Because RNA molecules are so simple, and phenotypes will often be expressed through folding patterns, we expect low β to be common in the RNA world. Therefore, we expect complete cooperation to be rare amongst RNAs.

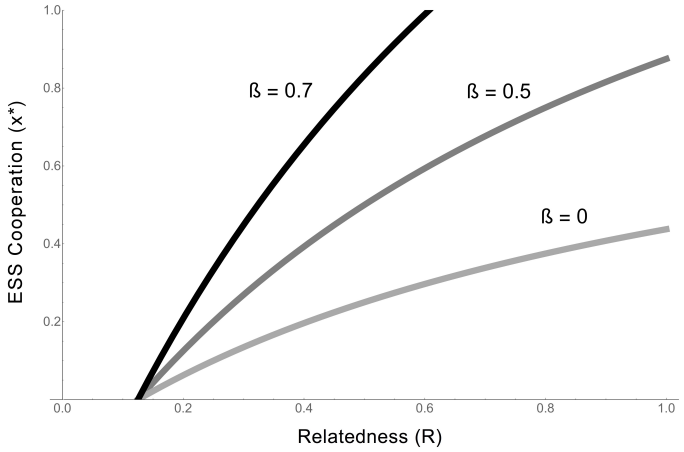

Figure 5. Do cooperators receive the benefits of cooperation? The y-axis shows the ESS level of cooperation (x^*), plotted against relatedness (R). The three lines represent three different values of β, which measures the degree to which cooperators receive benefits. When β is high, cooperators are equally as likely to receive the benefits of cooperative acts as non-cooperators. When β is low, acting as a cooperative RNA limits or prevents a molecule from receiving the benefits of cooperation. When β is low, only partial ($x^* << 1$) cooperation can evolve. For all values of β, increasing relatedness (R) increases the ESS value of cooperation. For all lines $b = 0.8$, $c = 0.1$.

2.3. Cooperation and Competition in RNA World

We have shown above that positive genetic relatedness can help favour cooperation (Equations (3), (6), and (7)). However, we have implicitly assumed that relatives can be together for cooperative

interactions without also competing with each other. This might not always be the case. RNA molecules may compete for resources, like nucleotides or binding sites on a surface. Whether or not this selects against cooperation can depend on biochemical details.

For example, limited diffusion of molecules (e.g., in [13,16,46]) leads to relatives being near each other (high relatedness). However, it also leads to the individuals with which a replicator competes also being relatives. In that case, competition is relatively *local*, as the extra individuals produced by competition impact the local group. Alternatively, replicators might act locally to increase each other's replicative rates, but disperse in a vesicle to compete *globally*. For example, abiotic protocells formed from amphilic molecules could divide by shearing and combine with each other [47]. In this case, the extra copies produced by cooperation displace individuals at a global level. Exactly how do relatedness and competition interact and impact cooperation?

We can answer this question by extending the previous model (Equations (4)), taking the case of $\beta = 0$ for simplicity) to incorporate competition, with a new parameter, $0 \leq a \leq 1$. a measures the scale of competition, with a proportion a of competition occurring in the local social group, and the remaining $1 - a$ occurring globally [22,34,48]. For example, if $a = 0$, all competition occurs at the population level, with each individual competing equally with every other individual. If $a = 1$, all competition occurs within the local social group—extra offspring produced only displace local individuals. Because we are now distinguishing between local and global competition, we must distinguish between the average phenotype of the social group (y) and the average phenotype of the population ($\bar{y}$):

$$\begin{aligned} w(x, y, \bar{y}) &= \frac{f}{aF + (1-a)\mathbb{F}} \\ &= \frac{x(1-c) + (1-x)(1+by)}{a(y(1-c) + (1-y)(1+by)) + (1-a)(\bar{y}(1-c) + (1-\bar{y})(1+b\bar{y}))}. \end{aligned} \tag{8}$$

The term in the numerator is a focal replicator's fitness (f), which is a relative denominator that measures the average fitness the focal individual competes against. The denominator is composed of the local (F) and global ($\mathbb{F}$) average fitnesses, where the relative importance of each is determined by a.

The fitness function in Equations (8) is analogous to Takeuchi et al.'s model of the origin of genome-like molecules, which looked at the evolution of template-like, selfish molecules, and protein-like cooperative molecules from a single starting point (though we only capture the evolutionary, not ecological aspects of their model) [24]. In their model, replicators that act as cooperative catalysts (protein-like molecules) cannot also act as templates (DNA-like molecules) $\beta = 0$, and mutations vary the probability with which a replicator acts as one or the other.

The Taylor–Frank method gives the ESS to be

$$x^* = \frac{Rb - c - a(Rb - Rc)}{b(1+R) - a2Rb}. \tag{9}$$

Equation (9) tells us that:

1. When competition is completely global ($a = 0$), the ESS reverts to the that identified in the previous model (Equation (7), $\beta = 0$), and the condition for cooperation to evolve is simply $Rb - c > 0$. This confirms our previous results, and is a qualitatively similar result to that found by Takeuchi et al., as in their model competition is relatively global and replicators cooperate about half the time (Equation (7), [24]).
2. As competition becomes more local (a increases), the condition for cooperation to evolve becomes more stringent. When competition (a) is high, the competitive effects of the extra offspring copies produced by, e.g., cooperative enzyme activity, are experienced locally. This means that in addition to benefiting from cooperation, relatives suffer increased competition from cooperation, making it harder for cooperation to evolve (Figure 6).

3. When competition is completely local ($a = 1$), $x^* = -\frac{c}{b}$. Since the costs and benefits are both positive, cooperation cannot evolve.

We expect many RNA biologies to lead to local competition. This is because most RNA population structures that have been described involve simple diffusion, which does not afford many opportunities for exporting the benefits of cooperation globally. If this is the case, we should be on the lookout for either: (1) different forms of RNA dispersal, other than simple limited diffusion, which could decrease local competition; or (2) other features of RNA biology that might achieve the same effect. We provide an example of the latter in the next section.

The above analysis also links RNA world with the wider literature on the influence of local competition. Previous work on bacteria and insects has demonstrated how local competition can reduce selection for cooperation [39,49,50]. Furthermore, local competition can also select for harmful or spiteful traits that reduce the fitness of competitors [48,51–53]. We might expect to find such traits in the RNA world.

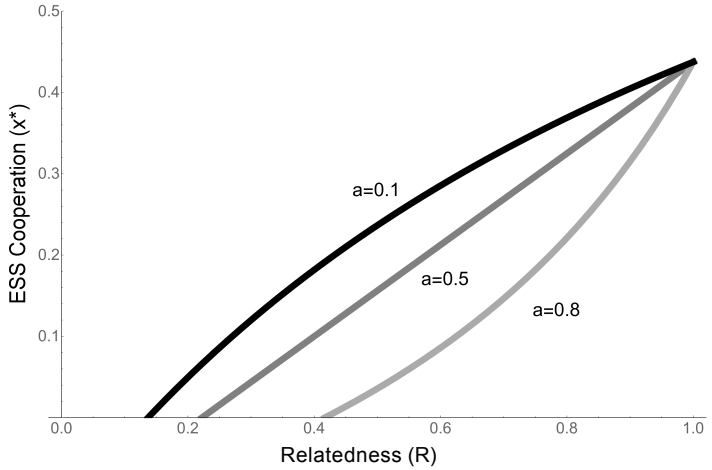

Figure 6. The scale of competition in the RNA world. The y-axis shows the ESS, x^*, derived from the model in the text for different parameter values. The x-axis is relatedness (R). The three lines represent three different values of a, the scale of competition. When a is low, competition is relatively global. When it is high, competition is relatively local. Increasing a reduces the ESS value of cooperation. For all values of a, increasing relatedness favours cooperation. For all lines $b = 0.8$, $c = 0.1$.

2.4. Explicit Population Structure: Closing the Model

In our analysis above we implicitly assumed that R is independent from the other model parameters. In reality this is unlikely to be the case. For example, limited diffusion leads to both local competition *and* higher relatedness, and so a and r should be positively correlated [34,35,54,55].

We can allow for this by modelling an explicit RNA life history, the parameters of which can then be used to calculate an estimate of relatedness in system. Incorporating specific life history parameters and using them to calculate relatedness in the model is called 'closing the model'. This is an alternative to our previous models, which assumed independence between R and model parameters ('open' models) [54,56]. We develop an explicit population structure model known as an infinite island model. In an island model, an infinite population is subdivided into patches, or groups of individuals, of size N. Island models are standard in evolutionary biology, but distinct from the lattice approach often used in RNA models, in that we do not explicitly track distance. While this island model approach is taken for analytical tractability, it has been shown to give qualitatively similar results to explicit lattice structures [57].

We extend the model of cooperation in Equation (5) by explicitly incorporating two life history parameters: offspring diffusion and parent survival. We expect both parameters to impact relatedness

and competition. Offspring remaining nearby, and parents remaining between generations both increase relatedness, but to different degrees, also impacting the scale of competition. An explicit model can determine the full impact of each of these factors.

We assume parent RNA molecules produce offspring copies, and then a proportion of parents, k, survives into the next generation, and the remaining $1-k$ fraction of parents dies. After reproduction, a proportion, d, of offspring copies diffuses elsewhere, while the other $1-d$ fractions remain locally. An individual's fitness can now be decomposed into the probability it survives and has fitness 1, or dies, and therefore its fitness is a function of whether its offspring diffuse or remain. If offspring diffuse they compete globally, and if they remain they compete locally, against the patch average fitness. We can write fitness as a function of the focal individual's phenotype, x, the average phenotype of the other individuals on the patch, y, and as the whole-group average (including the focal individual), Z ($\frac{y(N-1)+x}{N}$) (see Supplementary Materials for derivation):

$$\begin{aligned} w(x,y,Z) = (1-k)(d)(x(1-c+by)+(1-x)(1+by))+ \\ (1-d)\frac{x(1-c+by)+(1-x)(1+by)}{1+(1-d)(bZ(1-Z)-cZ)}+k. \end{aligned} \tag{10}$$

Note that the RNA system captured by the fitness function in Equation (10) is analogous to Shay et al.'s model of trans-replicases, in which two complementary strands of a trans-replicase can act as (selfish) templates or (cooperative) replicases, which increase the replication rate of templates [16,17]. We do not explicitly track the different phenotypes of the complements, instead looking at the overall probability a replicator is a cooperator (e.g., replicase) or selfish (e.g., template), and looking at the evolution of this probability.

In the Supplementary Materials, we use the Taylor–Frank method to identify the ESS in terms of the model parameters and R. We then calculate an estimate of the equilibrium value of R in terms of the model parameters by writing a recursion for how population parameters change R from one generation to the next. Once the equilibrium value of R is determined, we substitute back in for R to get an ESS of

$$x^* = \frac{2bk(1-d)-c(2k(1-d)+N(2+k-d-k(1-d)))}{b(1+k)N-b(1-d)(k(N-4)-N)}. \tag{11}$$

Equation (11) is a result we previously attained in a model of RNA trans-replicases [17]. It tells us that:

1. Cooperation is favoured by higher parent survival (larger k), limited diffusion (lower d), and smaller local group size (smaller N).
2. Both parent survival ($k > 0$) and limited diffusion ($d < 1$) are required for cooperation to evolve. While discrete generations ($k = 0$) hold approximately for many higher organisms, we expect overlapping generations ($k > 0$) to hold for most RNA systems, because we expect RNAs to survive well after their copies' copies have replicated. This offers a potential explanation for why RNA molecules might have cooperated despite having a dispersal strategy that otherwise leads to high local competition.

The approach we have used in this section is a closed modelling approach, where the relatedness emerged from the population parameters of the model (diffusion, survival, etc.), rather than being an open parameter. The benefit of this approach is that it reveals the exact relationship between model parameters and cooperation. The downside is that it required being explicit about the life history and population structure of the RNA molecules. In situations in which we know these details this approach will be useful. Otherwise it may be useful to keep the model open, and subsume unknown population processes in R.

3. Discussion

We have used the analytical methods of kin selection to model cooperation in the RNA world. We started with a deliberately simple model, abstracting away biochemical details to identify a general process by which cooperation is favoured in the RNA world. We showed that cooperation in RNAs, like many other organisms, is favoured by positive genetic relatedness (our R) (Equation (3)). Genetic relatedness can arise a number of ways, such as active kin discrimination, or just limited dispersal (population viscosity). Previous explanations for RNA cooperation include limited diffusion, spiral waves, and primitive cells, all of which serve to generate high genetic relatedness (Table 1). We have shown that these previously disparate explanations can be unified under the single explanatory framework of kin selection.

We elaborated on our most simple model by incorporating specific biological details that might be especially relevant in RNA world. We examined whether cooperative RNAs can receive the benefits of cooperation (Equation (4)). We expect that it will be common for RNA cooperators to be unable to receive the full benefits of cooperation, which suggests that complete cooperation will be rare in the RNA world (Figure 5) [4,16,17,24]. We then modelled the scale of competition in the RNA world (Equation (8)). The simple RNA population structures that generate high relatedness, which favours cooperation, are also likely to lead to high local competition, which we have shown disfavours RNA cooperation (Figure 6) (Table 1). This suggests that other life history features are likely to be important for overcoming local competition [17]. We explored this possibility in our final model ((Equation (10)), by examining whether cooperation can evolve under different conditions of offspring diffusion and parent survival. We showed that for a simple RNA system, both limited diffusion and parent survival (overlapping generations) are necessary for the evolution of cooperation ((Equation (11)).

We have made a number of assumptions that are standard in kin selection analyses, and it is worth questioning their validity for the RNA world. One issue is that we have assumed that all phenotypes in the strategy set (e.g., $0 \leq x \leq 1$) are possible. With very simple RNAs this may not be the case, as the phenotype space may not be continuous. In this case, we would need to limit the strategy set, for example to a number of discrete strategies [30,58]. Another issue is that we have assumed that near-equilibrium viable mutant variants have mutations of small effect (weak selection) [21,33]. However, again, the simple nature of RNA molecules may mean that mutations tend to be of large effect [4]. In this case an explicit population genetic model could be more appropriate. Whether these two assumptions are borne out, and how our predictions would change if they were violated, remains to be investigated both theoretically and empirically.

Table 1. Applying the kin selection approach to example RNA world systems. Rows show different example RNA world model systems, as well as references that have modelled such systems. The columns show four questions which can be asked of a model system: How is relatedness generated? Do cooperators receive benefits? What is the scale of competition? How can the predictions made by such models be tested? Subcolumns (e.g., limited diffusion, comparative) give the answers to those questions. The X symbol shows which of the subcolumns apply to the model system in a given row.

Example Model System	Relatedness Is Generated by			Do Benefits Return to Cooperators? (β)		Scale of Competition (a)		Testable Predictions?	
	Limited Diffusion	Proto-Cells	Other Spatial Clustering	Yes (High β)	No (Low β)	Global (Low a)	Local (High a)	Comparative	Experimental
Replicases in protocells [5,9,15]		X		X		X			X
Replicases on surfaces [12,13]	X			X			X		X
Trans-replicases [16]	X				X		X	X	X
Nucleatase and polymerase [46]	X			X			X	X	X
Origin of genome-like molecules [24]			X	X			X		X
Hyper-cycles [8,11]			X	X			X		X

3.1. Why Bother?

We have shown that we can think of various types of cooperation in the RNA world as being driven by kin selection, and that we can use social evolution tools to model this process. However, an obvious question arises: If one can also model these processes using other tools such as simulations, why bother with the kin selection methodologies that we have used here? We suggest three main benefits.

3.1.1. Generality

First, the simple analytical nature of these models offers biological insight and generally applicable conclusions. For example, we generated models that focused on the effects of dispersal, cooperation type, or the scale of competition. This approach lends itself to systems in which specific biochemical details are obscured, as is the case in the RNA world, because we do not yet know what the first replicators looked like. Streamlined models isolate key parameter relationships and identify important general biological features, which allows us to extend our conclusions beyond specific systems. Further, we have found that our models make similar predictions to more explicit simulations (e.g., Equation (9) and [24] or Equation (11) and [16]), which means that what we gain in generality is not necessarily lost in predictive power. Finally, when very different approaches lead to the same predictions, we gain confidence in those predictions.

3.1.2. Testability

Second, kin selection has been useful for creating a link between theory and experiments in higher organisms, and it should offer the same for the RNA world [20,59]. Kin selection models often identify simple relationships between parameters and traits, that can be tested with both experiments and across-species comparative studies (e.g., [36–42,59–61]).

For example, Figure 5 could be tested through comparative work. Different replicators synthetised in the lab will have different conformational properties: some will be more or else readable by a polymerase when folded to act as a cooperative enzyme, or will be more else able to fold and unfold and therefore express different phenotypes. Equation (7) makes a simple, testible prediction about the amount of cooperation we expect to see in these different replicators. Figure 6 could be tested experimentally in the lab (experimental evolution), by manipulating the scale of competition, as has been done with bacteria [39]. For example, solutions of RNA replicators could be well mixed (global competition, low relatedness), growing on surfaces (local competition, high relatedness), or growing on surfaces but migrating distantly on the surface of beads (high relatedness, global competition), and Equation (9) predicts the level of cooperation we expect to evolve under each of these conditions. More generally, this approach identifies relatedness as an important parameter to manipulate experimentally [39,41,49,50,62]. These are a handful of examples, but the link to kin selection provides a wealth of empirical examples to draw from (see Table 1), and simple parameters to test comparatively or experimentally.

Further, kin selection provides a simple heuristic to frame our thinking, generate verbal predictions, and identify fruitful evolutionary problems in the RNA world, which are all advantages for the empiricist. Take, for example, the evolution of a replicase enzyme which copies template strands (e.g., in [16]). We do not have to think of this as cooperation being driven by kin selection. However, doing so identifies life history features that are likely to be important (e.g., diffusion and survival), points us to problems that may arise that have been well studied in other taxa (e.g., the link between competition and relatedness), and makes it easy to generate predictions that can be tested in the lab (a system of replicators with limited diffusion and parent survival should evolve cooperation, whereas a system with limited diffusion alone should not). As origin of life experimental capabilities become more advanced, kin selection could become an increasingly useful tool for guiding empirical work and making testable predictions, as has been the case with animals and microbes [19,20,26,59,63,64].

3.1.3. Conceptual Links

Finally, the kin selection approach allows us make conceptual links to other taxa, unifying our evolutionary explanations across the tree of life. The tools we have used here are the same as those used to study cooperation in bacteria, birds, and mammals [59,61], and we can use them to identify common factors favouring cooperation across taxa. One advantage of this approach is that that we can use insight from the vast existing kin selection literature to guide our thinking about the RNA world. For example, our model of competition (Equation (8)) identified spite as a potentially important trait to consider in the RNA world. This approach also has the advantage of simplifying our understanding of life, providing a unifying framework rather than generating new explanations for each case. This last advantage is a key goal of science. A number of simulation studies have already demonstrated the utility of more complex, explicit approaches to solving problems in the RNA world (reviewed by [7]). We are not saying that the RNA world must be conceptualised using kin selection, just that it can be useful to do so.

Supplementary Materials: The following are available online at www.mdpi.com/2075-1729/7/4/53/s1, Supplementary Materials: Additional Mathematical Derivations.

Acknowledgments: We thank Paul Higgs for inviting us to contribute to this special issue, Geoff Wild, Guy Cooper, Matishalin Patel, and two anonymous reviewers for helpful comments and feedback, and the Clarendon Fund, Hertford College, and Natural Environment Research Council for funding.

Author Contributions: S.R.L. and S.A.W. contributed equally to the work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RNA	Ribonucleic acid
ESS	Evolutionarily stable strategy

References

1. Crick, F.H. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367–379.
2. Orgel, L.E. Evolution of the genetic apparatus. *J. Mol. Biol.* **1968**, *38*, 381–393.
3. Gilbert, W. Origin of life: The RNA world. *Nature* **1986**, *319*, 618.
4. Eigen, M.; Schuster, P. A principle of natural self-organization. *Naturwissenschaften* **1977**, *64*, 541–565.
5. Szathmáry, E.; Demeter, L. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **1987**, *128*, 463–486.
6. Smith, J.M.; Szathmary, E. *The Major Transitions in Evolution*; Oxford University Press: Oxford, UK, 1995.
7. Higgs, P.G.; Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **2015**, *16*, 7–17.
8. Boerlijst, M.C.; Hogeweg, P. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Phys. D Nonlinear Phenom.* **1991**, *48*, 17–28.
9. Frank, S.A. Kin selection and virulence in the evolution of protocells and parasites. *Proc. R. Soc. Lond. B Biol. Sci.* **1994**, *258*, 153–161.
10. Boerlijst, M.C.; Hogeweg, P. Spatial gradients enhance persistence of hypercycles. *Phys. D Nonlinear Phenom.* **1995**, *88*, 29–39.
11. Cronhjort, M.B.; Blomberg, C. Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Phys. D Nonlinear Phenom.* **1997**, *101*, 289–298.
12. McCaskill, J.S.; Füchslin, R.M.; Altmeyer, S. The stochastic evolution of catalysts in spatially resolved molecular systems. *Biol. Chem.* **2001**, *382*, 1343–1363.
13. Szabó, P.; Scheuring, I.; Czárán, T.; Szathmáry, E. In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **2002**, *420*, 340–343.

14. Sardanyés, J.; Solé, R.V. Spatio-temporal dynamics in simple asymmetric hypercycles under weak parasitic coupling. *Phys. D Nonlinear Phenom.* **2007**, *231*, 116–129.
15. Bianconi, G.; Zhao, K.; Chen, I.A.; Nowak, M.A. Selection for replicases in protocells. *PLoS Comput. Biol.* **2013**, *9*, e1003051.
16. Shay, J.A.; Huynh, C.; Higgs, P.G. The origin and spread of a cooperative replicase in a prebiotic chemical system. *J. Theor. Biol.* **2015**, *364*, 249–259.
17. Levin, S.R.; West, S.A. The evolution of cooperation in simple molecular replicators. *Proc. R. Soc. Lond. B Biol. Sci.* **2017**, *284*. doi:10.1098/rspb.2017.1967.
18. West, S.A.; Griffin, A.S.; Gardner, A. Evolutionary explanations for cooperation. *Curr. Biol.* **2007**, *17*, R661–R672.
19. Bourke, A.F. *Principles of Social Evolution*; Oxford University Press: Oxford, UK, 2011.
20. Davies, N.; Krebs, J.; West, S. *An Introduction to Behavioural Ecology*, 4th ed.; Wiley-Blackwell: Hoboken, NJ, USA, 2012.
21. Hamilton, W.D. The genetical theory of social behavior. I and II. *J. Theor. Biol.* **1964**, *7*, 1–52.
22. Frank, S.A. *Foundations of Social Evolution*; Princeton University Press: Princeton, NJ, USA, 1998.
23. Nee, S. The evolutionary ecology of molecular replicators. *R. Soc. Open Sci.* **2016**, *3*, 160235.
24. Takeuchi, N.; Hogeweg, P.; Kaneko, K. The origin of a primordial genome through spontaneous symmetry breaking. *Nat. Commun.* **2017**, *8*, 250.
25. West, S.A.; Buckling, A. Cooperation, virulence and siderophore production in bacterial parasites. *Proc. R. Soc. Lond. B Biol. Sci.* **2003**, *270*, 37–44.
26. West, S.A.; Diggle, S.P.; Buckling, A.; Gardner, A.; Griffin, A.S. The social lives of microbes. *Annu. Rev. Ecol. Evol. Syst.* **2007**, *38*, 53–77.
27. Frank, S.A. A general model of the public goods dilemma. *J. Evolut. Biol.* **2010**, *23*, 1245–1250.
28. Gardner, A.; West, S.A.; Wild, G. The genetical theory of kin selection. *J. Evolut. Biol.* **2011**, *24*, 1020–1043.
29. Biernaskie, J.M.; West, S.A. Cooperation, clumping and the evolution of multicellularity. *Proc. R. Soc. B R. Soc.* **2015**, *282*, 1075.
30. Smith, J.M.; Price, G. The Logic of Animal Conflict. *Nature* **1973**, *246*, 15–18.
31. Taylor, P.D.; Frank, S.A. How to make a kin selection model. *J. Theor. Biol.* **1996**, *180*, 27–37.
32. Hamilton, W.D. Selfish and spiteful behaviour in an evolutionary model. *Nature* **1970**, *228*, 1218–1220.
33. Grafen, A. A geometric view of relatedness. *Oxf. Surv. Evolut. Biol.* **1985**, *2*, 28–89.
34. West, S.A.; Pen, I.; Griffin, A.S. Cooperation and competition between relatives. *Science* **2002**, *296*, 72–75.
35. Lehmann, L.; Rousset, F. How life history and demography promote or inhibit the evolution of helping behaviours. *Philos. Trans. R. Soc. B Biol. Sci.* **2010**, *365*, 2599–2617.
36. Hughes, W.O.; Oldroyd, B.P.; Beekman, M.; Ratnieks, F.L. Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **2008**, *320*, 1213–1216.
37. Cornwallis, C.; West, S.; Griffin, A. Routes to indirect fitness in cooperatively breeding vertebrates: Kin discrimination and limited dispersal. *J. Evolut. Biol.* **2009**, *22*, 2445–2457.
38. Cornwallis, C.K.; West, S.A.; Davis, K.E.; Griffin, A.S. Promiscuity and the evolutionary transition to complex societies. *Nature* **2010**, *466*, 969–972.
39. Griffin, A.S.; West, S.A.; Buckling, A. Cooperation and competition in pathogenic bacteria. *Nature* **2004**, *430*, 1024–1027.
40. Diggle, S.P.; Gardner, A.; West, S.A.; Griffin, A.S. Evolutionary theory of bacterial quorum sensing: When is a signal not a signal? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2007**, *362*, 1241–1249.
41. Kuzdzal-Fick, J.J.; Fox, S.A.; Strassmann, J.E.; Queller, D.C. High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. *Science* **2011**, *279*, 1548–1551.
42. Lukas, D.; Clutton-Brock, T. Cooperative breeding and monogamy in mammalian societies. *Proc. R. Soc. B R. Soc.* **2012**, *334*, 2151–2156.
43. West, S.A.; Griffin, A.S.; Gardner, A. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J. Evolut. Biol.* **2007**, *20*, 415–432.
44. Von der Dunk, S.H.; Colizzi, E.S.; Hogeweg, P. Evolutionary Conflict Leads to Innovation: Symmetry Breaking in a Spatial Model of RNA-Like Replicators. *Life* **2017**, *7*, 43.
45. Frank, S.A. Multivariate analysis of correlated selection and kin selection, with an ESS maximization method. *J. Theor. Biol.* **1997**, *189*, 307–316.
46. Kim, Y.E.; Higgs, P.G. Co-operation between Polymerases and Nucleotide Synthetases in the RNA World. *PLoS Comput. Biol.* **2016**, *12*, e1005161.

47. Zhu, T.F.; Szostak, J.W. Coupled growth and division of model protocell membranes. *J. Am. Chem. Soc.* **2009**, *131*, 5705–5713.
48. Gardner, A.; West, S.A. Spite and the scale of competition. *J. Evolut. Biol.* **2004**, *17*, 1195–1203.
49. West, S.A.; Murray, M.G.; Machado, C.A.; Griffin, A.S.; Herre, E.A. Testing Hamilton's rule with competition between relatives. *Nature* **2001**, *409*, 510.
50. Kümmerli, R.; Gardner, A.; West, S.A.; Griffin, A.S. Limited dispersal, budding dispersal, and cooperation: An experimental study. *Evolution* **2009**, *63*, 939–949.
51. Gardner, A.; West, S.A.; Buckling, A. Bacteriocins, spite and virulence. *Proc. R. Soc. B R. Soc.* **2004**, *271*, 1529–1535.
52. Gardner, A.; Hardy, I.C.; Taylor, P.D.; West, S.A. Spiteful soldiers and sex ratio conflict in polyembryonic parasitoid wasps. *Am. Nat.* **2007**, *169*, 519–533.
53. West, S.A.; Gardner, A. Altruism, spite, and greenbeards. *Science* **2010**, *327*, 1341–1344.
54. Taylor, P.D. Altruism in viscous populations—An inclusive fitness model. *Evolut. Ecol.* **1992**, *6*, 352–356.
55. Queller, D.C. Genetic relatedness in viscous populations. *Evolut. Ecol.* **1994**, *8*, 70–73.
56. Gardner, A.; West, S. Demography, altruism, and the benefits of budding. *J. Evolut. Biol.* **2006**, *19*, 1707–1716.
57. Comins, H.N.; Hamilton, W.D.; May, R.M. Evolutionarily stable dispersal strategies. *J. Theor. Biol.* **1980**, *82*, 205–230.
58. Grafen, A. The hawk-dove game played between relatives. *Anim. Behav.* **1979**, *27*, 905–907.
59. West, S. *Sex Allocation*; Princeton University Press: Princeton, NJ, USA, 2009.
60. Fisher, R.M.; Cornwallis, C.K.; West, S.A. Group formation, relatedness, and the evolution of multicellularity. *Curr. Biol.* **2013**, *23*, 1120–1125.
61. Bourke, A.F. Hamilton's rule and the causes of social evolution. *Phil. Trans. R. Soc. B* **2014**, *369*, 20130362.
62. Frank, S.A. Hierarchical selection theory and sex ratios. II. On applying the theory, and a test with fig wasps. *Evolution* **1985**, *39*, 949–964.
63. West, S.A.; Griffin, A.S.; Gardner, A.; Diggle, S.P. Social evolution theory for microorganisms. *Nat. Rev. Microbiol.* **2006**, *4*, 597–607.
64. Wild, G.; Gardner, A.; West, S.A. Adaptation and the evolution of parasite virulence in a connected world. *Nature* **2009**, *459*, 983.

4

The social coevolution hypothesis for the origin of the genome

The social coevolution hypothesis for the origin of the genome

Abstract

At the start of life, the origin of the genome required individual replicators, or genes, to act like enzymes and cooperatively copy each other. The evolutionary stability of such enzymatic cooperation poses a problem, because it would have been susceptible to parasitic replicators, that don't act like enzymes, but could still benefit from the enzymatic behaviour of other replicators. Existing hypotheses to solve this problem require restrictive assumptions that may not be justified, such as the evolution of a cell membrane before the evolution of a genome, or very particular patterns of diffusion on special types of surfaces. We show theoretically that, instead, selection itself can lead to replicators grouping themselves together in a way that favours cooperation. We show that the tendency to physically associate to others and cooperative enzymatic activity can coevolve, leading to the evolution of physically linked cooperative replicators. Our results shift the empirical problem from a search for special environmental conditions to questions about what types of phenotypes can be produced by simple replicators.

1 Introduction

Even the simplest genomes are made up of hundreds of genes and thousands of base-pairs, and yet, by necessity, life started with single, short sequences, or replicators. Lacking the ability to produce large enzymes, these first replicators would have had high error rates in replication, preventing them from elongating into a genome (Eigen, 1971; Eigen and Schuster, 1977). There is a significant gap between the maximum size replicators could reach without large error-correcting enzymes, and the minimum size needed to produce those enzymes (Eigen and Schuster, 1977). Consequently, to bridge this gap and build a genome, different replicators would have had to act as enzymes to help copy each other (Eigen and Schuster, 1977; Smith and Szathmary, 1995). In this way, the individual replicators could remain small and below the error threshold, but the collection of replicators could grow sufficiently large to produce big enzymes.

The problem is that a collection of cooperative replicators would have been susceptible to parasitic replicators that don't act as enzymes, but are able to benefit from the enzymatic activity of others (Smith, 1979). All else being

equal, such molecular parasites (cheats) would have had a higher replication rate, making cooperation between replicators unstable, and hence preventing the evolution of a genome. What, then, can explain the maintenance of the cooperative enzymatic activity required for the genome to evolve? One hypothesis is that different types of replicators were grouped together in a primitive cell, or proto-cell, so that selection acted on groups of replicators (Smith and Szathmary, 1995; Szathmáry and Demeter, 1987; Frank, 1994; Bianconi et al., 2013). An alternative hypothesis is that replicators were on some surface which limited their diffusion, but also led to interactions between different types of cooperative replicators (Boerlijst and Hogeweg, 1991, 1995; Cronhjort and Blomberg, 1997; McCaskill et al., 2001; Szabó et al., 2002; Sardanyés and Solé, 2007; Shay et al., 2015). Both of these hypotheses favour cooperation by grouping cooperative replicators together, and hence limiting the extent to which they can be exploited by parasites.

However, these hypotheses require restrictive assumptions that may not be justified. In order to have replicators grouped by a cell membrane, we would require the evolution of a cell membrane before we had a genome that was sufficiently complex to produce that membrane. This solves the problem of explaining one complex feature (cooperative replicators) by invoking another complex feature (cell membrane). The proto-cell could be an abiotic feature, such as a droplet of oil, but that would require that the division of that droplet was linked to the rate at which replicators copy, in a way that just happened to make group selection work. The limited diffusion hypothesis requires evolution on a particular type of surface to group replicators together in a very specific way, which: (i) limits diffusion, so that parasites cannot exploit replicators; (ii) has high enough diffusion to keep different types of replicators well-mixed; (iii) has some special property which ensures binding sites contain different types of replicators, rather than copies of identical replicators (Szabó et al., 2002; Shay et al., 2015). It is not clear how a surface could produce all these properties. In addition, many previous models require simple replicators to have conditional phenotypes, and only act as cooperators in certain interactions, which is a relatively complex behaviour for a very short sequence (Shay et al., 2015; Levin and West, 2017).

We propose an alternative hypothesis, where selection itself leads to replicators grouping themselves together in a way that favours cooperative enzymatic activity. We hypothesise a scenario, where: (i) one type of replicator can evolve to act as an enzyme to help copy other replicators (cooperation); and (ii) another type of replicator can evolve to physically associate with or 'stick' to other replicators. We show theoretically that coevolution between these two traits can lead to cooperation between replicators being evolutionarily stable, in conditions where it would not otherwise be favoured. This occurs because the evolution of physical association allows the benefits of enzymatic cooperation to return to cooperators and their identical copies. This relatively simple hypothesis does not require restrictive features of the environment to group replicators together in certain ways, or the evolution of another complex feature of life, such as a cell membrane. Instead, selection drives the replicators to solve the problem

of cooperation. Consequently, our hypothesis shifts empirical focus from restrictive external environmental conditions to the biology of simple replicators themselves.

2 Model and Results

2.1 The life cycle

We imagine two different replicators, X and Y, which are independent populations but can form XY complexes, where a complex is an interacting pair of replicators. For simplicity, we do not track XX and YY pairings, as we assume that these pairings do not affect replication rate (in the appendix we show that a model that explicitly tracks these pairings leads to similar conclusions). These replicators could be RNA-like molecules, but the model is not limited to the RNA-world hypothesis. The only requirement is that the molecules are able to self-replicate (they are 'autocatalytic'), and can potentially act as catalysts for the replication of others (they possess 'enzyme-like' behaviour). We make no explicit assumptions about population structure except that individuals may occasionally interact and these interactions may affect their replication rate (fitness). Consequently, our model could apply to replicators interacting on a surface or free-floating. The biological interactions we envisage are depicted in Figure 1.

We can consider the population of replicators as divided into three populations: X replicators on their own, Y replicators on their own, and XY complexes. The densities of these populations are free to grow and shrink, and these densities affect the rates at which different interactions occur. The X and Y replicators each have some baseline, potentially distinct rates of replication ($\rho_{i\in X,Y}$) and destruction, or death ($\mu_{i\in X,Y}$). These two types of replicator form complexes with each other at some low baseline rate (β), and these complexes dissociate at some (relatively high) baseline rate (δ), or else are ended by the destruction of one of the replicators (notation summarised in Table 1).

When in complexes, replicators produce new individual replicators at a rate ($\theta_{i\in X,Y}$), which we assume to be higher than their rate of replication when on their own. This could be due to a beneficial waste product, like a nucleotide, produced during replication, or a conformational change passively induced by the other replicator which increases the efficiency of replication. This byproduct benefit is measured by κ, such that $\theta_{i\in X,Y} = (1+\kappa)\rho_{i\in X,Y}$. We also assume that the new replicators produced by complexes can immediately pair again, due, for example, to proximity (r_{XY}). We imagine that, initially, this happens very rarely (though this is not required). In the appendix, we show that the population dynamics of these three different populations are described by:

$$\begin{aligned}
\frac{d[X]}{dt} &= (\rho_X - \mu_X)\,[X] - \beta[X][Y] + (\mu_Y + \theta_X + \delta)\,[XY] \\
\frac{d[Y]}{dt} &= (\rho_Y - \mu_Y)\,[Y] - \beta[Y][X] + (\mu_X + \theta_Y + \delta)\,[XY] \\
\frac{d[XY]}{dt} &= \beta[X][Y] - (\mu_X + \mu_Y + \delta - r_{XY})\,[XY]
\end{aligned} \tag{1}$$

2.2 Evolutionary dynamics

We use an adaptive dynamics approach to study the evolution of cooperative enzymatic activity and physical associations in these replicators (Metz et al., 1992; Rand et al., 1994; Geritz et al., 1997; Dieckmann and Law, 1996; Van Baalen and Jansen, 2001). To do so, we follow three steps. First, we consider a mutant whose cooperative enzymatic activity or tendency to associate and dissociate differs from the resident population. Second, we determine what direction these traits will evolve in, by studying the spread of mutants (given by the initial asymptotic growth rate of a mutant with deviant trait values, or invasion fitness). Third, by allowing for successive mutants, we determine numerically the evolutionarily stable resting state of the population (Smith and Price, 1973).

We show in the appendix that condition for the spread of a rare mutant in replicator X or Y ($i \in X, Y$) can be expressed as

$$\frac{F_i'}{\beta'\overline{[j]}} + \frac{P_i'}{M_{ij}'} > 1. \tag{2}$$

$F_i = (\rho_i - \mu_i)$ is the replication rate of replicator $i \in X, Y$ on its own, $P_i = (\mu_j + \theta_i + \delta)$ is the replication rate of $i \in X, Y$ in complexes, and $M_{ij} = (\mu_i + \mu_j + \delta - \rho_{ij})$ is loss (destruction or dissociation) of complexes. The primes indicate mutant values in replicator $i \in X, Y$, and mutants are denoted $i' \in X', Y'$. Equation 2 shows how a trait can spread via its effect on the replication rate of a replicator on its own (F_i'), the effect on its replication rate in pairs (P_i'), the effect on the loss of complexes (M_{ij}'), and the effect on pairings with the other replicator type ($\beta'\overline{[j]}$) (Van Baalen and Jansen, 2001, derived a similar expression for the invasion of a rare mutant). We now proceed to study the evolution of cooperative enzymatic activity and physical association.

2.3 Enzymatic cooperation

We first asked whether selection would favour replicators to act as enzymes that help copy other replicators. This can be thought of as evolution towards more cooperative replicators, which would facilitate the evolution of the genome. We examined this possibility by allowing replicator X to mutate in a way that made it better at helping copy replicator Y, by increasing the density independent replication of Y by a factor $1+\omega d'$. We assumed that this mutation would cause

the X replicator to be less efficient at copying itself, by reducing the replication rate of X by a factor $1 - d'$. For example, this could be a conformational change which reduces X's autocatalytic rate, but causes X to act as a catalyst to increase the replication rate of Y. Consequently, we are assuming a trade-off between the rate at which a replicator can help copy other replicators, and the rate at which that replicator can copy itself.

Replicator copies produced from complexes may immediately form pairs again, and it is possible that, through increasing the local density of Y replicators, an X' mutant increases the chance that its copies immediately pair again with a Y. To account for this, we assume an X' mutant increases the rate that replicators produced from complexes immediately find a partner by a factor $1 + \lambda d'$ (where λ might equal ω but is free to vary). In the appendix, we extend the model to explicitly track this effect, and recover similar results.

We found that cooperative enzymatic activity was not favoured. Specifically, more cooperative X' mutants ($d' > 0$) were never able to invade a population of resident X and Y, and that the X population rests stably at a value of zero cooperation (Figure 2a). We found cooperation could not spread because it reduced the replication rate of the mutant, and there was no mechanism by which the benefit to Y could be fed back to X'. While cooperative enzymatic activity increases the density of Y, the baseline association rates are sufficiently low that this effect is not strong enough to favour such activity. Specifically, cooperation reduces both terms in equation 2, by reducing both replication in complexes and alone (the numerators), and leaving the association with the other type ($\beta'\overline{[j]}$) unaffected. Cooperation reduces the loss of complexes in the second term (M'_{ij}), but this is not enough to outweigh the direct cost to replication. This is analogous to the standard evolutionary result that, all else being equal, a cooperative behaviour that benefits an unrelated individual will not be favoured (Hamilton, 1964).

2.4 Physical association

We then examined the consequences of allowing the Y replicator to mutate in a way that causes it to associate or form complexes with the X replicator, increasing the baseline association rate (β) by a factor $1 + \zeta c'$, and decreasing the rate at which complexes dissociate (δ) by a factor $1 - \xi c'$ (where $0 \geq \xi \leq 1$). We refer to this as an 'association' trait, as it can capture the possibility that Y' physically binds to X (e.g. 'stickiness'), but also includes any kind of trait which increases the rate of association between X and Y' and/or increases the duration of these associations, such as a trait which induced a conformational change in X increasing the chance they form a pair. We allow only mutations in Y, holding X constant.

We assume that this association mutation is costly, and decreases the rate at which Y' can replicate itself by a factor $1-c'$, for example because, due to its new folding pattern, it is less easily replicated. We account for the possibility that this association trait increases the chance that copies produced from complexes immediately pair again by allowing the mutation to increase the rate of pairing

by a factor $1 + \alpha c'$. A baseline assumption might be that $\alpha = \zeta$, because this effect is simply due to the increase in association rate caused by the association mutation, but the model allows for the effect to be weaker or stronger.

We found that association could be favoured (Figure 2b). Specifically, if the byproduct benefits gained by being paired with an X (κ), and the relative increase in association rate caused by the mutation ($\frac{\alpha c'}{c'}$) are sufficiently high ($>> 1$), successive mutations with higher values of association ($c' > 0$) will invade a population of resident X and Y, until the association rate comes to rest at some equilibrium value ($c*$). Some level of association is favoured because while it causes an immediate reduction in replication rate, this is outweighed by the increase in replication rate due to being in complexes with X more often. Specifically, association reduces the first term in equation 2 (via the numerator), but this is outweighed by an increase in the second term (via the denominator).

2.5 Coevolution

We then examined what happens when both enzymatic cooperation and association are allowed to co-evolve. We did this by allowing for mutations in both replicators: X to evolve to be more cooperative ($d' > 0$), and Y to associate at a greater rate ($c' > 0$). To allow for coevolution, we analysed the selection gradient on both traits in both mutant populations simultaneously, which tells us which direction in state space the population is moving at any given point. By repeating this across all of state space for both traits, we can determine which direction both replicator types will evolve in.

In the appendix we show that, when both traits are allowed to coevolve, selection can drive enzymatic cooperation ($d*$) to its maximal value and association ($c*$) to a higher value than when evolving on its own (Fig. 2c). Coevolution favours enzymatic cooperation when the association and enzymatic cooperation increase the chance that replicators produced by complexes pair again ($\lambda, \alpha >> 1$), and when byproduct benefits are large ($\kappa >> 1$). Further, even under conditions in which association would not evolve on its own (e.g. when $\zeta = 0$), if association still increases the duration of pairings ($\xi > 0$), then coevolution can favour enzymatic cooperation and association.

This result is driven by coevolution between the two traits. In the absence of the association trait, cooperation is not favoured because the benefits only accrue to members of the other replicator type. But as association evolves, there is an increased chance that a cooperator's copies both form and remain in pairs with an associator mutant's copies. A cooperator increases the chance that its copies will find a Y partner by creating more Y copies.

Our results show that the key factor favouring positive co-evolution between cooperation and physical association is that the two traits increase the chance that copies produced from complexes pair again. We captured this in the term $(1+\lambda d')(1+\alpha c'))\rho_{ij}$, which leaves unspecified exactly how the two traits increase immediate pairing. In the appendix, we derive an explicit model, tracking the individual copies produced from pairs, and modelling how they repair. The explicit model recovers the results of the more general model, showing that

co-evolution is only favoured when both traits increase the chance of pairing again.

More generally, cooperation between different replicators is conceptually analogous to cooperation between different species in mutualisms. The X and Y populations of replicators can be thought of as two different 'species'. Cooperation can be favoured between species when the benefits of cooperation return to the co-operator or its genetic relatives (Frank, 1998; Foster and Wenseleers, 2006; Gardner et al., 2006; Wyatt et al., 2013). In our model, the interaction between the traits provides a mechanism for the benefits of cooperation to return to the cooperator's copies, and the association trait prevents these relationships from breaking down. This link can be made formally with a multi-locus population genetic model of replicator evolution, where cooperation is driven by the combination of physical association and selection on the fitness of cooperator pairs (Supplementary Material).

Another force which has been shown to drive such positive between-species co-evolution is synergy of fitness effects (Gardner et al., 2006; Queller, 2011). Synergy occurs when two co-operators together do better than expected because the whole is greater than the sum of the parts. In the appendix, we show, using a multi-locus model, that synergy favours cooperation. While we did not include this in our explicit ecological model, we expect that the addition of synergistic effects would further favour the evolution of cooperation.

3 Conclusions

We proposed and tested a hypothesis that the coevolution between enzymatic activity and physical associations can explain cooperation between different types of replicators. We showed that if one population of replicators can act as an enzyme to increase the replication rate of another, and the other can act to increase the physical associations between the two, these traits can coevolve, given there is some baseline byproduct benefit to being complexes. This leads to a population of replicators, or genes, which are both cooperative and physically linked, the two key features of a genome. Specifically, in our scenario the questions of why simple replicators would come together physically (byproduct benefits) and how they would overcome the error threshold (cooperation) resolve each other.

Our results make the evolution of a primitive genome easier to explain, by simplifying the conditions required. This does not mean that previous explanations are invalid, just that they may come in at different stages in the evolution of life. For the genome, our results suggest that we don't need to: (a) invoke the cell, a potentially complex feature of life; (b) assume highly specific population structures on special surfaces; or (c) grant simple replicators with conditional phenotypes. Consequently, our result increases the kinds of environments where the genome can evolve. It also means that a more complex genome could have evolved to then produce the first cell, because our result shows how genome complexity could increase without a cell membrane. Finally, our results shift

the focus of origin of the genome questions from external features of the environment to biological features of replicators. Specifically, what phenotypes are possible in simple molecules?

References

Barton, N. and Turelli, M. (1991). Natural and sexual selection on many loci. *Genetics*, 127(1):229–255.

Bianconi, G., Zhao, K., Chen, I. A., and Nowak, M. A. (2013). Selection for replicases in protocells. *PLoS Comput Biol*, 9(5):e1003051.

Boerlijst, M. C. and Hogeweg, P. (1991). Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D: Nonlinear Phenomena*, 48(1):17–28.

Boerlijst, M. C. and Hogeweg, P. (1995). Spatial gradients enhance persistence of hypercycles. *Physica D: Nonlinear Phenomena*, 88(1):29–39.

Cronhjort, M. B. and Blomberg, C. (1997). Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Physica D: Nonlinear Phenomena*, 101(3-4):289–298.

Dieckmann, U. and Law, R. (1996). The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of mathematical biology*, 34(5-6):579–612.

Eigen, M. (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523.

Eigen, M. and Schuster, P. (1977). A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565.

Foster, K. R. and Wenseleers, T. (2006). A general model for the evolution of mutualisms. *Journal of evolutionary biology*, 19(4):1283–1293.

Frank, S. A. (1994). Kin selection and virulence in the evolution of protocells and parasites. *Proceedings of the Royal Society of London B: Biological Sciences*, 258(1352):153–161.

Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.

Gardner, A., West, S. A., and Barton, N. H. (2006). The relation between multilocus population genetics and social evolution theory. *The American Naturalist*, 169(2):207–226.

Geritz, S. A., Metz, J. A., Kisdi, É., and Meszéna, G. (1997). Dynamics of adaptation and evolutionary branching. *Physical Review Letters*, 78(10):2024.

Hamilton, W. D. (1964). The genetical theory of social behavior. i and ii. *Journal of Theoretical Biology*, 7(1):1–52.

Hurford, A., Cownden, D., and Day, T. (2009). Next-generation tools for evolutionary invasion analyses. *Journal of the Royal Society Interface*, 7(45):561–571.

Kirkpatrick, M., Johnson, T., and Barton, N. (2002). General models of multilocus evolution. *Genetics*, 161(4):1727–1750.

Levin, S. R. and West, S. A. (2017). The evolution of cooperation in simple molecular replicators. *Proceedings of the Royal Society of London B: Biological Sciences*, 284(1864).

McCaskill, J. S., Füchslin, R. M., and Altmeyer, S. (2001). The stochastic evolution of catalysts in spatially resolved molecular systems. *Biological chemistry*, 382(9):1343–1363.

Metz, J. A., Nisbet, R. M., and Geritz, S. A. (1992). How should we define âĂŸfitnessâĂŹ for general ecological scenarios? *Trends in Ecology & Evolution*, 7(6):198–202.

Queller, D. C. (2011). Expanded social fitness and hamilton's rule for kin, kith, and kind. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10792–10799.

Rand, D. A., Wilson, H., and McGlade, J. M. (1994). Dynamics and evolution: evolutionarily stable attractors, invasion exponents and phenotype dynamics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 343(1305):261–283.

Sardanyés, J. and Solé, R. V. (2007). Spatio-temporal dynamics in simple asymmetric hypercycles under weak parasitic coupling. *Physica D: Nonlinear Phenomena*, 231(2):116–129.

Shay, J. A., Huynh, C., and Higgs, P. G. (2015). The origin and spread of a cooperative replicase in a prebiotic chemical system. *Journal of theoretical biology*, 364:249–259.

Smith, J. M. (1979). Hypercycles and the origin of life. *Nature*, 280:445–446.

Smith, J. M. and Price, G. (1973). The logic of animal conflict. *Nature*, 246(5427):15–18.

Smith, J. M. and Szathmary, E. (1995). *The major transitions in evolution*. Oxford University Press.

Szabó, P., Scheuring, I., Czárán, T., and Szathmáry, E. (2002). In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature*, 420(6913):340–343.

Szathmáry, E. and Demeter, L. (1987). Group selection of early replicators and the origin of life. *Journal of theoretical biology*, 128(4):463–486.

Van Baalen, M. and Jansen, V. A. (2001). Dangerous liaisons: the ecology of private interest and common good. *Oikos*, 95(2):211–224.

Wyatt, G., West, S., and Gardner, A. (2013). Can natural selection favour altruism between species? *Journal of evolutionary biology*, 26(9):1854–1865.

4 Appendix A

4.1 The model

We need a total of seven equations to capture the ecological dynamics of both the residents and the mutants. From the life cycle in figure one and the description in the main text, this allows us to write (for the residents):

$$\begin{aligned}
\frac{d[X]}{dt} =& \left(\left(r_X\left(1-d\right)\eta\right)-\mu_X\right)[X]\right) \\
&-\left(\beta\left(1+\zeta c\right)[X][Y]\right) \\
&+\left(\left(\left(1+\kappa\right)\left(1-d\right)r_X\eta\right)+\mu_Y+\left(1-\xi c\right)\delta\right)[XY] \\
\frac{d[Y]}{dt} =& \left(\left(r_Y\left(1-c\right)\eta\right)-\mu_Y\right)[Y]\right) \\
&-\left(\beta\left(1+\zeta c\right)[Y][X]\right) \\
&+\left(\left(\left(1+\kappa\right)r_Y\left(1-c\right)\left(1+\omega d\right)\eta\right)+\mu_X+\left(1-\xi c\right)\delta\right)[XY] \\
\frac{d[XY]}{dt} =& \beta\left(1+\zeta c\right)[Y][X] \\
&-\left(\mu_Y+\mu_X+\left(1-\xi c\right)\delta-\left(\left(1+\lambda d\right)\left(1+\alpha c\right)\right)r_{XY}\eta\right)[XY]
\end{aligned}$$

Where $\eta = 1 - k([T] = [X] + [XY] + [Y])$ is density dependent regulation, and k controls its extent.

Following the standard adaptive dynamics approach, we assume that invading mutants are rare enough so as not to affect the dynamics of the resident population (Metz et al., 1992; Rand et al., 1994; Dieckmann and Law, 1996). Accordingly, we only need four additional equations to capture the dynamics of mutants in each gene (X' and Y'), which are expressions for $\frac{d[X']}{dt}$, $\frac{d[Y']}{dt}$, $\frac{d[X'Y]}{dt}$, and $\frac{d[XY']}{dt}$. These differ from system 3 only in their values for c and d, the mutant trait values, which we denote with primes as c' and d'.

The equations for the mutant can be written in the form (Van Baalen and Jansen, 2001):

$$\frac{d[i']}{dt} = F_i'[i'] - \beta'[i']\overline{[j]} + P_i'[i'j] \tag{3}$$

$$\frac{d[i'j]}{dt} = \beta'[i']\overline{[j]} - M_{ij}'[i'j] \tag{4}$$

Where $F_i = (\rho_i - \mu_i)$ is the growth of i alone, $P_i = (\mu_j + \theta_i + \delta)$ is the production of i's from complexes, and $M_{ij} = (\mu_i + \mu_j + \delta - \rho_{ij})$ is the loss of complexes. The primes indicate mutant values in gene i. This form for an invasion condition was first identified by Van Baalen and Jansen (2001). For illustration, we reproduce each term for replicator Y, but equivalent equations can be extracted for replicator X.

$$\begin{aligned} F_Y' &= (r_Y (1 - c') ((1 - k([T])))) - \mu_Y \\ P_Y' &= ((1 + \kappa) r_Y (1 - c') (1 + \omega d) ((1 - k([T]))) + \mu_X + (1 - \xi c') \delta) \\ M_{XY}' &= (\mu_Y + \mu_X + (1 - \xi c') \delta - ((1 + \lambda d) (1 + \alpha c')) r_{XY} ((1 - k([T])))) \\ \beta_Y &= \beta (1 + \zeta c') \end{aligned} \tag{5}$$

We can rewrite system 5 in matrix form as:

$$\frac{d}{dt}\begin{bmatrix}[i'] \\ [i'j]\end{bmatrix} = \begin{bmatrix} F_i' - \beta'\overline{[j]} & P_i' \\ \beta'\overline{[j]} & -M_{ij}' \end{bmatrix}\begin{bmatrix}[i'] \\ [i'j]\end{bmatrix} \tag{6}$$

The first matrix on the right hand side of equation 6 contains all the information we need to determine the spread of a rare mutant (Van Baalen and Jansen, 2001; Hurford et al., 2009). A useful decomposition of this matrix is the form $\mathbf{F_i}$ - $\mathbf{V_i}$, where,

$$\mathbf{F_i} = \begin{bmatrix} F_i' & P_i' \\ 0 & 0 \end{bmatrix}, \mathbf{V_i} = \begin{bmatrix} \beta_i'\overline{[j]} & 0 \\ -\beta_i'\overline{[j]} & M_{ij}' \end{bmatrix} \tag{7}$$

According to the next generation theorem (Hurford et al., 2009), given that $\mathbf{F} > 0$, $\mathbf{V^{-1}}$, and the spectral bound of $-\mathbf{V}$ is negative, the condition for a mutant to invade is that the spectral radius of $\mathbf{FV}^{-1} > 1$. This condition can be written as:

$$\frac{P_i'}{M_{ij}'} + \frac{F_i'}{\beta'\overline{[j]}} > 1 \tag{8}$$

This is equation 2 in the main text.

Taking the derivatives of equation 8 with respect to small changes in mutant trait values gives the direction of selection in each trait, which we use to produce Figure 2c.

Coevolution is driven by the interaction between the two mutant trait values, c' and d'. These interaction terms are contained entirely in the derivative of the first term of Equation (8, and it can easily be seen that $\frac{d}{di}\left(\frac{P_i'}{M_{ij}'}\right)$ is increasing with respect to changes in c' and d'.

4.2 Tracking same-type replicator pairs

Above we did not track XX or YY pairs. This means that the above model holds in systems where the replicators do not form self-self complexes. We also conjectured that the results would approximately hold even if they do form such complexes, because individuals in XX and YY pairs do not gain byproduct benefits, and therefore have a lower replication rate than when in XY complexes. We checked this by developing a model that explicitly tracks such pairings. This requires two additional equations for the density of XX and YY complexes, for a total of five equations:

$$\begin{aligned}
\frac{d[X]}{dt} =& \left(\left(r_X\left(1-d\right)\eta\right)-\mu_X\right)[X]\right) \\
&-\left(\beta\left(1+\zeta c\right)[X][Y]\right) \\
&\quad+\left(\left(\left(1+\kappa\right)\left(1-d\right)r_X\eta\right)+\mu_Y+\left(1-\xi c\right)\delta\right)[XY] \\
&\quad-\beta[X][X] \\
&\quad+\left(\mu_X+\delta+r_X\left(1-d\right)\eta\right)[XX] \qquad (9) \\
\frac{d[Y]}{dt} =& \left(\left(r_Y\left(1-c\right)\eta\right)-\mu_Y\right)[Y]\right) \\
&-\left(\beta\left(1+\zeta c\right)[Y][X]\right) \\
&\quad+\left(\left(\left(1+\kappa\right)r_Y\left(1-c\right)\left(1+\omega d\right)\eta\right)+\mu_X+\left(1-\xi c\right)\delta\right)[XY] \\
&\quad-\beta[Y][Y] \\
&\quad+\left(\mu_Y+\delta+r_Y\left(1-d\right)\eta\right)[YY] \qquad (10) \\
\frac{d[XY]}{dt} =& \beta\left(1+\zeta c\right)[Y][X] \\
&-\left(\mu_Y+\mu_X+\left(1-\xi c\right)\delta-\left(\left(1+\lambda d\right)\left(1+\alpha c\right)\right)r_{XY}\left(1-k\left([T]\right)\right)\right)[XY] \\
\frac{d[XX]}{dt} =& \beta[X][X]-\left(\mu_X+\mu_Y+\delta-r_{XX}\eta\right)[XX] \\
\frac{d[YY]}{dt} =& \beta[Y][Y]-\left(\mu_Y+\mu_X+\delta-r_{YY}\eta\right)[YY]
\end{aligned}$$

Where $\eta = 1-k([X]+[XY]+[Y]+[XX]+[YY])$.

Following the same approach as above, we derive the condition for a mutant in replicator type i to spread as:

$$\frac{\overline{[i]}\beta}{\overline{[i]}\beta+\overline{[j]}\beta'}\left(\frac{P'_{ii}}{M'_{ii}}\right)+\frac{\overline{[j]}\beta'}{\overline{[i]}\beta+\overline{[j]}\beta'}\left(\frac{P'_i}{M'_{ij}}+\frac{F'_i}{\overline{[j]}\beta'}\right)>1 \qquad (11)$$

The new terms, P'_{ii} and M'_{ii}, capture the production and loss of same type pairs, respectively. This inequality is of a similar form to Equation (8). The original expression for fitness is now weighted by the relative rate of pairing with the other type. The new component of fitness (the first term) the ratio of

production of same type pairs to loss of same type pairs, and is weighted by the relative rate of pairing with the same type.

Numerically solving across parameter state space shows that the same results hold as above, with cooperative enzymatic activity failing to spread on its own, association evolving in the absence of such activity, and the two traits co-evolving to higher values than when on their own (Supplementary Material).

4.3 An explicit model of pairing

The above model left unspecified how cooperation and association increase pairing of replicator copies, capturing the effect in the term ρ_{ij}. We now adapt the model to a specific population structure, in order to make this effect explicit. Doing so necessarily requires sacrificing some of the generality of the first model, but what it loses in generality it gains in precision.

We now need to track two additional populations: X's and Y's that have been produced from pairs. This is because in order to explicitly model the effects of cooperation and association on pairing, we need to track the densities of copies produced from pairs before they become randomly mixed in the population.

The assumptions are the same as above, except now we allow for some baseline rate, χ, at which copies produced from pairs immediately pair again. Otherwise they return to the independent populations of X and Y. We assume that the rate of pairing is increased by both cooperation and physical association, by a factor $(1+\lambda d')(1+\alpha c')$, and is a function of the densities of copies produced, denoted $[X_o]$ and $[Y_o]$. The new system of equations describing the population dynamics is now:

$$\begin{aligned}
\frac{d[X]}{dt} &= (((r_X(1-d)\eta) - \mu_X)[X]) \\
&\quad - (\beta(1+\zeta c)[X][Y]) + \psi[X_o] \\
\frac{d[Y]}{dt} &= (((r_Y(1-c)\eta) - \mu_Y)[Y]) \\
&\quad - (\beta(1+\zeta c)[Y][X]) + \psi[Y_o] \\
\frac{d[XY]}{dt} &= \beta(1+\zeta c)[Y][X] \\
&\quad - (\mu_Y + \mu_b + (1-\xi c)\delta)[XY] + (1+\lambda d)(1+\alpha c')[X_o][Y_o] \\
\frac{d[X_o]}{dt} &= ((1+\kappa)(1-d) r_X \eta)[XY] - (1+\lambda d)(1+\alpha c')[X_o][Y_o] - \psi[X_o] \\
\frac{d[Y_o]}{dt} &= ((1+\kappa) r_Y (1-c)(1+\omega d)\eta)[XY] \\
&\quad - (1+\lambda d)(1+\alpha c')[X_o][Y_o] - \psi[Y_o], \qquad (12)
\end{aligned}$$

where $\eta = 1 - k([X] + [XY] + [Y] + [X_o] + [Y_o])$. The parameter ψ controls the relative rate at which copies produced from pairs return to the population of free X and Y.

Following the same approach as before, we can write system 12 in matrix form as:

$$\frac{d}{dt}\begin{bmatrix}[i'] \\ [i'j] \\ [i'_o]\end{bmatrix} = \begin{bmatrix}F'_i - \beta'\overline{[j]} & P'_i & \psi \\ \beta'\overline{[j]} & -M'_{ij} & A'_i\overline{[j_o]} \\ 0 & PR'_i & -A'_i\overline{[j_o]} - \psi\end{bmatrix}\begin{bmatrix}[i'] \\ [i'j] \\ [i'_o]\end{bmatrix} \tag{13}$$

$P'i$ now measures only replicators returned to the independent population from complexes as a result of dissociation and destruction, because copies produced from complexes are captured in the term PR'_i. A'_i measures the association of copies produced from complexes, and $\overline{[j_o]}$ is the equilibrium frequency copies produced from complexes.

Using the next generation theorem (Hurford et al., 2009), we find the condition for a mutant to invade a resident population to be:

$$\frac{\psi P'_i + \overline{[j_o]}P'_iA'_i + \psi PR'_i}{\psi M'_{ij} + \overline{[j_o]}M'_{ij}A'_i - \overline{[j_o]}PR'_iA'_i} + \frac{F'_i}{\beta'\overline{[j]}} > 1 \tag{14}$$

This equation and its derivatives with respect to changes in c' and d' allow us to analyse evolution of cooperation or physical association on their own, or their co-evolution. We recover the result that co-evolution can favour the evolution of cooperation in conditions under which it would not have evolved on its own (Supplementary Material). The result depends crucially on the relative rate at which copies produced from pairs return to the independent populations (ψ), with $\psi \approx 1$ recovering the main results.

5 Supplementary Material

The model in the main text used an adaptive dynamics approach, which allowed us to explicitly track all of the ecological components of our system. However, the complexity required for such realism meant that we sacrificed the ability to develop simple analytical solutions. Such analytical solutions can help illuminate links to other problems of between-species cooperation across the tree of life.

Here, we develop an idealised model of cooperation between two replicator types, keeping it as simple as possible to gain interpretable, analytical results, and to help frame and understand the results from the adaptive dynamics model. We use a multilocus methodology adopted from population genetics (Barton and Turelli, 1991; Kirkpatrick et al., 2002; Gardner et al., 2006).

We imagine two replicators, X and Y. We assume that individuals interact in between-type pairs, such that, effectively, at any given time, the population is made up entirely of XY pairs. Finally, we imagine that each replicator can acquire a mutation which causes it to act as a cooperator. Cooperators increase the replication rate of their partner in the pair, at a cost (in terms of replication rate) to themselves. This could, for example, involve acting as an enzyme which

increases the replication rate of others, but, as a result of folding, decreases the cooperator's ability to self replicate.

We denote the strategy of an individual in replicator i as C_i for cooperate and D_i for defect. A cooperator in replicator i gives a benefit, b_i, to its partner, at a cost, c_i to itself. Defectors give no benefits and incur no costs. An individual's replication rate is a function of its own strategy and the strategy of its partner. We allow for some degree of synergy, or epistasis, by allowing for individuals in CC pairs to gain an additional, synergistic effect, d, which we assume is the same for both replicators, and can be positive or negative. For example, two interacting enzymes could enhance ($d > 0$) or inhibit ($d < 0$) each other's activity.

We can separately track the two effects of selection: selection at the level of pairs (which favours cooperators) and selection at the level of replicators (which favours defectors) (Gardner et al., 2006). To do this, we assume the fitness of a pair is the average of the fitness of its replicators. In DD pairs, the fitness of the individual replicators is identical, and so the group fitness measure is an actual count of the number of 'offspring' pairs produced. In all other pairs, there will be excess members of one replicator type produced and a deficit of another (in any given pair). To account for this, we assume that excess individuals of each gene replace missing members in other groups. This requires that the frequency of the cooperation mutation is the same each replicator type. Thus, although we count the fitness of pairs, this is only to account for group selection – pairs don't actually replicate, only individuals do.

Finally, after replication, we allow some fraction, m, of pairs to dissociate and re-pair at random. This is sufficient to allow us to track the change in frequency of the cooperation mutation in both replicator types, and the change in association between individuals. All else being equal, we expect costly cooperation to be unable to invade. But if selection generates associations between cooperators, this may allow cooperation to evolve, because the cost of cooperation can be outweighed by the statistical association between being a cooperator and receiving cooperation in return. This is an extension of a previous model of between-species cooperation by Gardner et al. (2006), where here replicators, or genes, are individuals and we allow cooperators from each replicator type to have different fitness payoffs. We proceed to ask under what conditions a rare mutant cooperator in either or both replicator types can invade.

We write X_i as the genetic value of cooperation in replicator i, with $X = 0$ for defection and $X = 1$ for cooperation. From the payoffs given in figure one and the description of the life cycle in the main text, we can write the fitness, W, of a pair of replicators as:

$$W = X_i X_j \left(\frac{2 - c - s + k + b + 2d}{2} \right) + (1 - X_i) X_j \left(\frac{2 - s + k}{2} \right) + X_i (1 - X_j) \left(\frac{2 - c + b}{2} \right) + (1 - X_i)(1 - X_j)(1) \tag{15}$$

Using standard multilocus techniques (Barton and Turelli, 1991; Kirkpatrick

et al., 2002; Gardner et al., 2006), we can extract selection coefficients from Equation (15). Writing $\mathfrak{a}_i$ for the selection coefficient of cooperation in replicator i, we get:

$$\mathfrak{a}_i = \frac{\frac{b-c}{2} + dp}{\overline{w}} \tag{16}$$

$$\mathfrak{a}_j = \frac{\frac{k-s}{2} + dp}{\overline{w}} \tag{17}$$

$$\mathfrak{a}_{ij} = \frac{d}{\overline{w}}$$

where mean fitness across the populations is $\overline{w} = 1 + \frac{1}{2}p\left(b - c + k - s\right) + d\left(\mathfrak{D}_{ij} + p^2\right)$ and $\mathfrak{D}_{ij}$ is the association between genetic values for cooperation across all pairs of replicators in the population. This is sufficient to allow us to write change in frequency of cooperation in each replicator type and the change in association between the cooperators each time step (episode of selection). Writing p_i as the frequency of the cooperators in replicator i, and using primes to distinguish between time steps, we get:

$$\begin{aligned} p_i' =& p_i + \mathfrak{a}_i p_i q_i + \mathfrak{a}_j \mathfrak{D}_{ij} + \mathfrak{a}_{ij}\left(1 - 2p_i\right)\mathfrak{D}_{ij} \\ \mathfrak{D}_{ij}' =& \mathfrak{D}_{ij} + \mathfrak{a}_i\left(1 - 2p_i\right)\mathfrak{D}_{ij} + \mathfrak{a}_j\left(1 - 2p_j\right)\mathfrak{D}_{ij} \\ &+ \mathfrak{a}_{ij}\left(p_i q_i p_j q_j + \left(1 - 2p_i\right)\left(1 - 2p_j\right)\mathfrak{D}_{ij} - \mathfrak{D}_{ij}^2\right) \end{aligned} \tag{18}$$

System (18) accounts for the effects of group selection. Turning to selection of replicators, and utilising the life cycle description in the main text, we can write 7 equations for the change, within groups, between generations, that results from individual level selection. Writing $t_{i\leftarrow i}$ as the probability that individual n in replicator i in a given pair in generation $h+1$ came from individual n in time h, and is therefore a cooperator, and $t_{i\leftarrow k}$ as the probability that it did not, and is therefore a defector, we get:

$$\begin{aligned} t_{i\leftarrow i} &= 1 - X_i\left(1 - X_j\right)\left(\frac{b+c}{2+b-c}\right) \\ t_{i\leftarrow k} &= X_i\left(1 - X_j\right)\left(\frac{b+c}{2+b-c}\right) \\ t_{j\leftarrow j} &= 1 - \left(1 - X_i\right)X_j\left(\frac{k+s}{2+k-s}\right) \\ t_{j\leftarrow h} &= \left(1 - X_i\right)X_j\left(\frac{k+s}{2+k-s}\right) \\ t_{ij\leftarrow ij} &= 1 - \left(X_i\left(1 - X_j\right)\left(\frac{b+c}{2+b-c}\right) + \left(1 - X_i\right)X_j\left(\frac{k+s}{2+k-s}\right)\right) \\ t_{ij\leftarrow kj} &= X_i\left(1 - X_j\right)\left(\frac{b+c}{2+b-c}\right) \\ t_{ij\leftarrow ih} &= \left(1 - X_i\right)X_j\left(\frac{k+s}{2+k-s}\right) \end{aligned} \tag{19}$$

Extracting transmission coefficients (Barton and Turelli, 1991; Kirkpatrick et al., 2002; Gardner et al., 2006) from System (19), we can write the change in frequency and association due to replicator selection (within pairs) as,

$$\begin{aligned} p_i'' &= p_i' - \frac{b+c}{2+b-c}\left(q_j' p_i' q_i' - q_i' \mathfrak{D}_{ij}'\right) \\ p_j'' &= p_j' - \frac{k+s}{2+k-s}\left(q_i' p_j' q_j' - q_j' \mathfrak{D}_{ij}'\right) \\ \mathfrak{D}_{ij}'' &= \mathfrak{D}_{ij}' \end{aligned} \tag{20}$$

Finally, accounting for the change in association due to diffusion (which does not impact frequency), we write:

$$\mathfrak{D}_{ij}''' = (1-d)\,\mathfrak{D}_{ij}'' \tag{21}$$

We consider the invasion of a rare mutant, which allows us to linearise all recursions in p_i, p_j and $\mathfrak{D}_{ij}$. We can write the change matrix for the system as,

$$\begin{bmatrix} p_i''' \\ p_j''' \\ \mathfrak{D}_{ij}''' \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 & \alpha_3 \\ 0 & \alpha_5 & \alpha_6 \\ 0 & 0 & \alpha_9 \end{bmatrix} \begin{bmatrix} p_i \\ p_j \\ \mathfrak{D}_{ij} \end{bmatrix} \tag{22}$$

The condition for cooperation to spread is that at least one of the eigenvalues, which are the diagonal coefficients, is greater than 1. The eigenvalues are given by

$$\begin{aligned} \lambda_1 &= 1-c \\ \lambda_2 &= 1-s \\ \lambda_3 &= (1-d)\,\frac{2+b-c+k-s+d}{2} \end{aligned} \tag{23}$$

The condition for the first two eigenvalues to be greater than 1 is simply that cooperation is selfish (the fitness of a cooperator paired with a defector is greater than the fitness of a defector paired with defector). The third eigenvalue gives the condition for the spread of cooperation:

$$(1-m)\,\frac{2+b-c+k-s+d}{2} > 1. \tag{24}$$

The first term in brackets captures diffusion, with increasing diffusion disfavouring cooperation, and the second term is the fitness of a replicator pair that contains two cooperators. The term on the right-hand side of the inequality is the fitness of a pair containing two defectors. Thus, the condition for cooperation to evolve is that the fitness of a CC pair, weighted by the rate at which pairs stay together, is greater than the fitness of a DD pair. This suggests that two key forces in allowing cooperation to evolve are physical association

and selection, with increasing synergy (d) favouring cooperation. This result is analogous to that obtained by Gardner et al. (2006).

Equation (24) supports and helps frame our result in the adaptive dynamics model: physical association and statistical associations generated by selection can favour cooperation between unrelated types of replicators. Equation (24) is a simple, easily interpreted analytical result which emerged from a model which made highly restrictive assumptions, including: (i) requiring that the frequency of the cooperative trait was the same in both replicators, (ii) forcing members of each species together, effectively assuming that individuals were only ever present in between-species pairs, and (iii) invoking an exogenous feature of the system, in the parameter m, which kept pairs together between generations. Our adaptive dynamics model relaxed all of those assumptions, by allowing traits in each species to evolve independently, allowing individuals to exist outside of pairs, and allowing the physical associations between individuals to be an evolving trait, and came to an analogous conclusion.

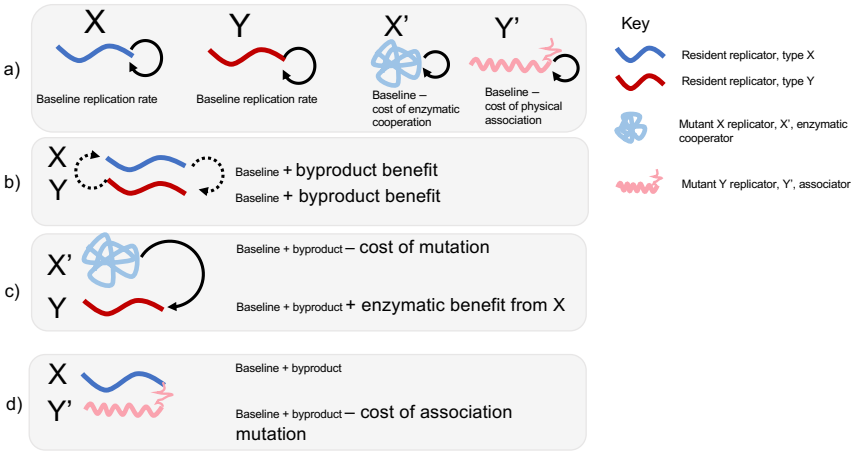

Figure 1: Interactions and fitness effects. (a) Two replicators, X and Y, each have some baseline replication rate on their own, but can acquire mutations (X' and Y') which reduce their replication rate. (b) Each replicator (X and Y) has a higher replication rate whenever in complexes due to passive benefits (byproduct benefits). (c) Mutant X' increases the replication rate of Y in complexes (enzymatic benefit from X). (d) Mutant Y' increasing the rate at which the mutant forms associations with X, and decreases the rate at which these associations break down.

Table 1: Summary of key notation

Notation	Definition
$[X]$	Density of replicator type X
$[Y]$	Density of replicator type Y
$[XY]$	Density of replicator pairs
ρ_i	Total production of type i replicators on their own
θ_i	Total production of type i replicators from pairs
r_i	Baseline replication rate of type i replicator
μ_i	Rate of destruction of type i replicators
β	Baseline association rate
δ	Baseline dissociation rate
k	Degree of density dependence
T	Total density of replicators in system
η	Density dependent replication, $= 1 - k[T]$
r_{ij}	Baseline rate at which ij pairs immediately pair again
κ	Byproduct benefit to being in pair
c	Degree of association trait
d	Degree of cooperative enzymatic activity trait
ζ	Increase in association rate due to association mutation
ξ	Decrease in dissociation rate due to association mutation
ω	Increase in replication rate of type Y due to cooperative enzymatic mutation
λ	Increase in the rate pairs re-pair due to cooperative enzymatic mutation
α	Increase in the rate pairs re-pair due to association mutation

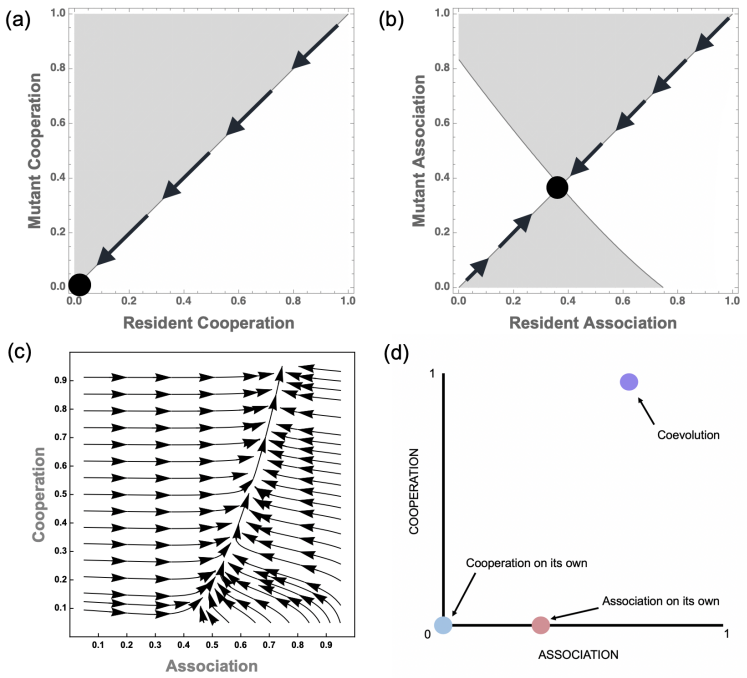

Figure 2: The coevolution of enzymatic activity and association. Grey shaded areas show regions of state space where selection is negative, and white areas where selection is positive. Arrows depict the direction of evolution in state space along neutral lines. Evolutionarily stable strategies are depicted by solid circles. (a) The evolution of enzymatic activity in X is not favoured. In the absence of association, cooperative enzymatic activity cannot evolve. (b) The evolution of association in Y. In the absence of cooperative enzymatic activity in X, some intermediate level of association in Y is favoured. (c) The coevolution of cooperative enzymatic activity and association. Arrows depict the direction of selection in both traits at a given point in state space. When traits are allowed to coevolve, cooperative enzymatic activity and association both evolve from anywhere in state space, with association reaching higher values than in (b), and enzymatic activity evolving towards its maximal value of 1. (d) A schematic of (a)-(c). Solid circles depict the evolutionarily stable resting point of both populations depending on whether each population evolves independently or evolve jointly. Coevolution favours higher values in both traits. Values for parameters in (a)-(d) are: $\alpha = 20, \lambda = 20, \zeta = 5, \xi = 1, \omega = 20, r_Y = 2.3, r_X = 2.1, r_{XY} = 0.9, k = 0.01, \mu_Y = 1.1, \mu_X = 1.1, \kappa = 100, \beta = 0.01, \delta = 0.9$. All figures generated graphically from the equations described in Appendix A using Mathematica Software version 11.3.0.0.

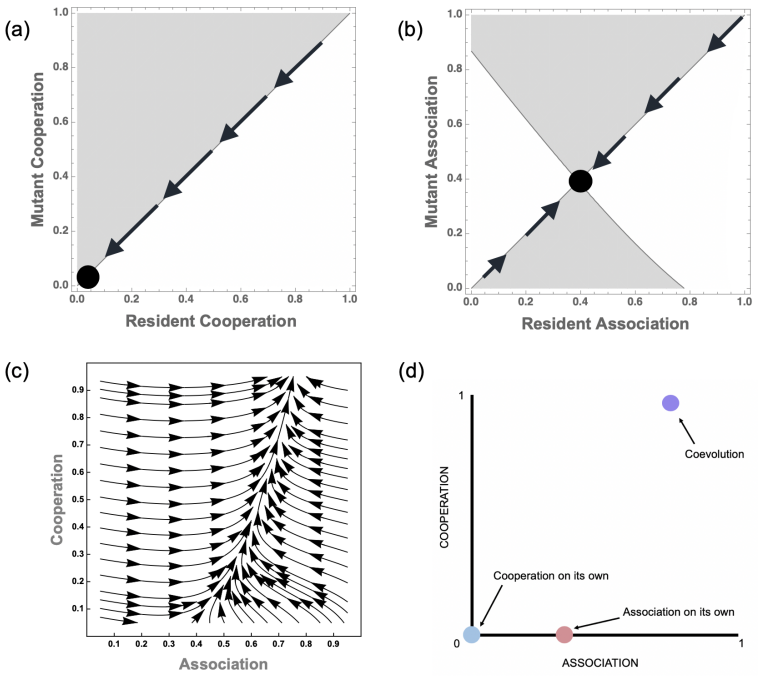

Figure S1: The coevolution of enzymatic activity and association when allowing for same-type pairs to form. Grey shaded areas show regions of state space where selection is negative, and white areas where selection is positive. Arrows depict the direction of evolution in state space along neutral lines. Evolutionarily stable strategies are depicted by solid circles. (a) The evolution of enzymatic activity in X is not favoured. In the absence of association, cooperative enzymatic activity cannot evolve. (b) The evolution of association in Y. In the absence of cooperative enzymatic activity in X, some intermediate level of association in Y is favoured. (c) The coevolution of cooperative enzymatic activity and association. Arrows depict the direction of selection in both traits at a given point in state space. When traits are allowed to coevolve, cooperative enzymatic activity and association both evolve from anywhere in state space, with association reaching higher values than in (b), and enzymatic activity evolving towards its maximal value of 1. (d) A schematic of (a)-(c). Solid circles depict the evolutionarily stable resting point of both populations depending on whether each population evolves independently or evolve jointly. Coevolution favours higher values in both traits. Values for parameters in (a)-(d) are: $\alpha = 20, \lambda = 20, \zeta = 5, \xi = 1, \omega = 20, r_Y = 2.3, r_X = 2.1, r_{XY} = 0.9, r_{XX} = 0.01, r_{YY} = 0.01, k = 0.01, \mu_Y = 1.1, \mu_X = 1.1, \kappa = 100, \beta = 0.01, \delta = 0.9$. All figures generated graphically from the equations described in Appendix A using Mathematica Software version 11.3.0.0.

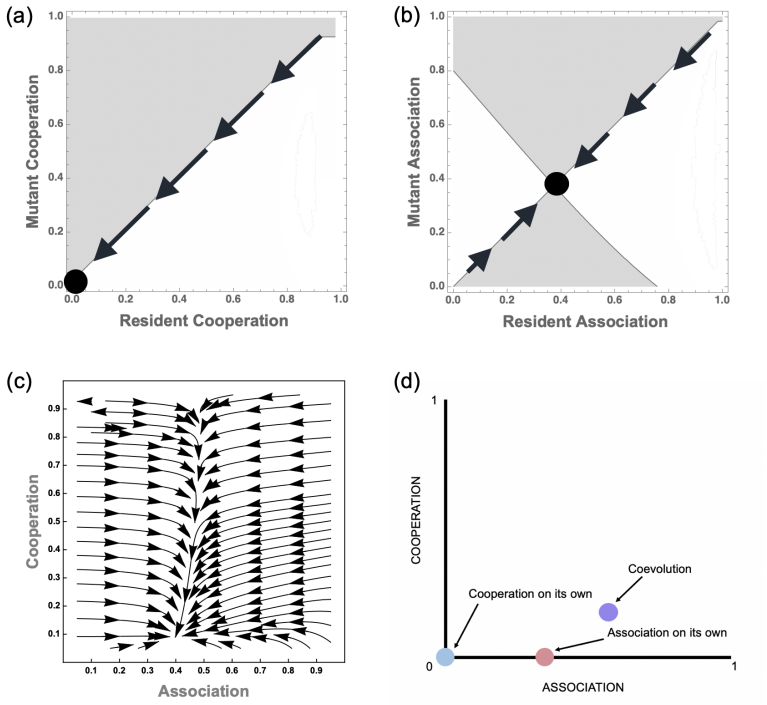

Figure S2: The coevolution of enzymatic activity and association in an explicit model of pairing. Grey shaded areas show regions of state space where selection is negative, and white areas where selection is positive. Arrows depict the direction of evolution in state space along neutral lines. Evolutionarily stable strategies are depicted by solid circles. (a) The evolution of enzymatic activity in X is not favoured. In the absence of association, cooperative enzymatic activity cannot evolve. (b) The evolution of association in Y. In the absence of cooperative enzymatic activity in X, some intermediate level of association in Y is favoured. (c) The coevolution of cooperative enzymatic activity and association. Arrows depict the direction of selection in both traits at a given point in state space. When traits are allowed to coevolve, cooperative enzymatic activity and association both evolve from anywhere in state space, with both traits reaching higher values than in (b) and (c). (d) A schematic of (a)-(c). Solid circles depict the evolutionarily stable resting point of both populations depending on whether each population evolves independently or evolve jointly. Coevolution favours higher values in both traits. Values for parameters in (a)-(d) are: $\alpha = 20, \lambda = 20, \zeta = 5, \xi = 1, \omega = 20, r_Y = 2.3, r_X = 2.1, r_{XY} = 0.9, k = 0.01, \mu_Y = 1.1, \mu_X = 1.1, \kappa = 100, \beta = 0.01, \delta = 0.9, \psi = 1$. All figures generated graphically from the equations described in Appendix A using Mathematica Software version 11.3.0.0.

5

Inclusive fitness is an indispensable approximation for understanding organismal design

Inclusive fitness is an indispensable approximation for understanding organismal design

Abstract

For some decades most biologists interested in design have agreed that natural selection leads to organisms acting as if they are maximising a quantity known as 'inclusive fitness'. This maximisation principle has been criticised on the (uncontested) grounds that other quantities, such as offspring number, predict gene frequency changes accurately in a wider range of mathematical models. Here we adopt a resolution offered by Birch, who accepts the technical difficulties of establishing inclusive fitness maximisation in a fully general model, while concluding that inclusive fitness is still useful as an organising framework. We set out in more detail why inclusive fitness is such a practical and powerful framework, and provide verbal and conceptual arguments for why social biology would be more or less impossible without it. We aim to help mathematicians understand why social biologists are content to use inclusive fitness despite its theoretical weaknesses. Here we also offer biologists practical advice for avoiding potential pitfalls.

Keywords: inclusive fitness, biological design, fitness maximisation, δ-weak selection, social evolution, population genetics

1. Introduction

Inclusive fitness was invented by Hamilton (1964) as an individual-level quantity (page 8) that natural selection should cause organisms to act as if maximising (page 17). The idea has been controversial for many decades (Cavalli-Sforza and Feldman, 1978) and there has been a recent explosion of controversy and debate (there are too many papers to cite here, but e.g. see Nowak et al., 2010 and replies, e.g. Abbot et al., 2011; Bourke, 2011; Queller, 2016) . We endorse and adopt here the resolution offered by Birch (2017a; 2017b), who accepts that the critics (e.g. Nowak et al., 2010; Allen and Nowak, 2016) are right to point to technical difficulties in establishing that inclusive fitness is well-defined *in a fully general theoretical model*, but at the same time concludes that the advocates (e.g. Grafen, 2006; Abbot et al., 2011; Gardner et al., 2011; West and Gardner, 2013; Queller, 2016; Marshall, 2015, 2016) have a strong enough case

within certain assumptions (notably additivity of fitness effects) to adopt inclusive fitness as an organising framework for understanding social behaviour. A goal here is to set out in more detail why inclusive fitness is such a practical and powerful organising framework, to such an extent that we argue the study of social behaviour would become more or less impossible without it.

In the course of the recent debate, several authors (e.g. West and Gardner, 2013; Queller, 2016) have written very clear arguments for some of the advantages of inclusive fitness, and readers are encouraged to refer to these papers for a general discussion of the role of inclusive fitness in biology (West and Gardner, 2013; Queller, 2016). However, our admittedly narrower focus here is to address mathematically rooted criticisms of the assumptions required to guarantee inclusive fitness maximisation, and the claim that measures such as mean-offspring number do a better job at predicting gene-frequency change. While this focus is narrower, it is also the controversial issue that continues to prevent productive dialogue between mathematicians and empiricists. Mathematical biologists making these points pay no regard to the practical arguments made by advocates of inclusive fitness, while still pointing to these formal shortcomings as a problem. Our goal here is to meet these mathematical arguments on their reasonable terms, and illustrate why, when interpreted in the light of whole-organism biology, many of the problems fall away.

To achieve this, we first outline five advantages of inclusive fitness. We initially focus on these advantages under additivity, to make the points clearly in the absence of the offending complications. We then turn to to the problem of non-additivity, and reconsider the advantages in this scenario. Finally, we discuss the importance of conditional behaviour in the degree to which non-additivity raises problems in practice, expanding on and clarifying points made previously (Grafen, 1979; Queller, 1996). We indicate how the necessary assumption of additivity can be checked in practical cases, and the likely impact of minor deviations. Anonymous, *Submitted* have shown formally that additive models are consistent with a very wide range of situations, and that inclusive fitness maximisation does occur in model circumstances where previous authors (Lehmann et al., 2015; Okasha and Martens, 2016b) have failed to find it. Here we focus on the less technical but broader conceptual arguments in support of those formal results, in a way that is accessible to non-mathematicians, and contains practical advice for empiricists.

2. Inclusive fitness under additivity

Hamilton (1964) observed that adult offspring number, a standard metric of fitness, is affected not just by the actions of an individual but by those of the individuals it interacts with. Hamilton pointed out that trying to measure those effects of relatives involves averaging over possible distributions of

genotypes, which in turn involves knowing gene frequencies in the population – a calculation he termed 'unwieldy' (Hamilton, 1964). However, he offered an alternative metric, which involves taking the perspective of the focal individual and its effects on others (as opposed to others' effects on it). He called this value 'inclusive fitness', and defined it as the sum of an individual's adult number of offspring in the absence of any social interactions (baseline fitness; more precisely, in the absence of social interactions in the performance of which there is genetic variability), and certain weighted effects the individual has on all individuals in the population, including itself. The effects are increases or decreases in offspring number caused by the individual, and the weightings are degrees of relatedness. Relatedness is a measure of genetic similarity between two individuals, with an individual having a relatedness of 1 to itself and 0 to a randomly selected member of the population. Inclusive fitness specifically does not include the effects of others on the focal individual.

Hamilton showed, under the assumption of weak selection, that this quantity, inclusive fitness, increases under selection, taking inspiration from Fisher's proof that standard fitness increases in an asocial model (Fisher, 1930; Hamilton, 1964, 1970), and modelling his technical argument on Kingman (1961). Hamilton argued that, as a result, we should expect organisms to appear as if trying to maximise their inclusive fitness (Hamilton, 1970). For nearly 40 years, at least within behavioural and evolutionary ecology, most field and laboratory workers have treated inclusive fitness as the quantity that organisms appear designed to maximise, and tailored their studies and experiments accordingly (summarised in, e.g. Westneat and Fox, 2010; Davies et al., 2012).

Inclusive fitness brings with it several advantages for the study of social behaviour. Here we outline five that we think are particularly important. In the following discussion, however, we are focusing on inclusive fitness under the assumption of additivity (though as we will see, all but the first extend beyond this restriction). There are two types of additivity. The first, 'additive gene action', is concerned with how different alleles combine within an individual to produce a phenotype, or social behaviour. Considering two alleles, A and B, is the difference between being AA and AB the same as that between AB and BB (additivity), or different (non-additivity)? The second type of additivity, which is of relevance here, refers to additivity of fitness effects between individuals. How does the effect of a social action combine with the existing number of offspring of an individual? And how do the effects of different social actions combine to affect one individual's offspring number? Let's say an individual has 5 offspring in the absence of social interactions, and social partners can choose to help that individual by giving it an extra 'b' offspring. Does each instance of helping simply add the same number onto the individual's existing number of offspring (additivity), or do those fitness effects combine in some

nonlinear way (non-additivity). Simple inclusive fitness models, which are used to make predictions about animal behaviour, assume additivity. We return to the problem of non-additivity later, as it is central to Birch's (2017a; 2017b) resolution of the debate.

2.1. Advantage 1: Predicting gene frequency change

The first advantage of inclusive fitness is that, under additivity, it correctly predicts the direction of gene frequency change. Hamilton's rule provides a simple tool for doing so (Hamilton, 1964). Given a trait that has an effect, in terms of adult offspring number, on its bearer, $-c$, and has an effect on social interactants, b, that trait will spread in the population if $rb - c > 0$, where r is the relatedness to the recipients affected by the trait. More generally, genes whose bearers tend to have a higher value of inclusive fitness will be favoured by natural selection (Hamilton, 1964, 1970). The rule easily extends to multiple recipients, though it is crucial that there is just one actor. Note that we are referring to the simple form of Hamilton's rule derived by Hamilton (1964), in which the fitness effects are absolute effects on offspring number, as this form is sufficient under non-additivity. We discuss the more general form (Queller, 1992; Gardner et al., 2011) in Section 5.2.

However, this advantage is rarely important on its own. It is the connection with other properties that makes predicting gene frequency change important in practice. We now go on to articulate those other properties.

2.2. Advantage 2: A design principle for individuals

Inclusive fitness provides a design principle for organisms. A fundamental question in biology (dating, in spirit if not detail, to Darwin) is how the dynamics of gene frequency change leads to the appearance of design and adaptation in organisms. Fisher's Fundamental Theorem (1930) provided such a link for non-social traits, by proving that natural selection always tends to increase mean fitness. It sometimes then follows that organisms appear designed as if trying to maximise that quantity. Hamilton established a similar result to Fisher's, but for social traits. Inclusive fitness is a quantity that, under additivity, organisms should appear designed to maximise (Hamilton, 1964; Rousset, 2015; Lehmann et al., 2016; West and Gardner, 2013; Lehmann and Rousset, 2014; Grafen, 2006; Gardner et al., 2011; Queller, 1992; Taylor, 2017).

Inclusive fitness is particularly useful as a design principle because it is can be conceptualised as an individual level property. Although it is possible to search for design principles at the level of the gene or the group, students of behaviour tend to predict and measure organismal phenotypes. The selfish gene approach can be useful for certain gene level questions, such as intragenomic conflict (Haig, 2002; Burt and Trivers, 2006; Foster, 2011; Gardner and Úbeda, 2017), whereas group level principles have been less useful (West et al., 2007, 2008;

Gardner and Grafen, 2009; West and Gardner, 2013). Individual level principles are the default tool of the trade (Davies et al., 2012), and have, in part, been successful because different loci in the genome tend to be selected in the same direction, and genetic rebels tend to be silenced by the 'parliament of the genes' (Leigh, 1977; Alexander and Borgia, 1978; Strassmann and Queller, 2010; West and Gardner, 2013). As a result, the different tissues and organs within an individual work together for a common cause, the good of the majority group of genes, which for shorthand we often call the good of the organism (Leigh, 1977; Haig, 2002; Burt and Trivers, 2006; Strassmann and Queller, 2010; West and Gardner, 2013).

If there is an individual level design principle in biology, then, at equilibrium, organisms should look like rational actors choosing, amongst a suite of available phenotypes, the one that maximises a certain quantity (Okasha and Martens, 2016b). Hamilton showed that, within his assumptions, there was such a quantity – inclusive fitness.

2.3. Advantage 3: Interpreting behaviour

Inclusive fitness provides a simple, economic interpretation of organismal behaviour (Hamilton, 1970; Grafen, 1984; Frank, 1998). Organisms should trade off their own offspring against those of another individual at a rate r (relatedness). This serves three purposes.

First, it helps generate testable predictions, even without complex mathematical models. Simple verbal reasoning can lead us to predict how many eggs a certain species of bird should lay each year, or how much food a cub should leave for its sibling, and these predictions are then readily testable.

Second, it guides us to new study systems by suggesting what biological features might lead to problematic or interesting cases. A heuristic for generating predictions is exactly how a scientific field makes progress, as has been demonstrated in the fields of behavioural ecology and evolutionary ecology (Krebs and Davies, 1978, 1987; Charnov, 1982; Krebs and Davies, 2009; West, 2009; Westneat and Fox, 2010; Davies et al., 2012).

Third, it helps us understand social behaviour by providing a way to reason about adaptations. For example, it's true that populations should be made up of genes that are associated with a higher contribution of gene copies to the next generation. But this doesn't tell us much about what kinds of traits and real life observations would defy our expectations, what population structures might lead to particularly unusual phenomena, or what adaptations (underpinned by many genes) might spread. Inclusive fitness offers us all of those things, by telling us that organisms should make decisions using this simple tradeoff in offspring.

2.4. Advantage 4: Empirical testability

An additional benefit of this simple tradeoff is that inclusive fitness predictions are testable in the laboratory and the field. Inclusive fitness, remarkably, doesn't require knowing the genetics of a trait (the 'phenotypic gambit'), the genotypes of various individuals in the population, or even gene frequencies (Grafen, 1984; West and Gardner, 2013). We only need to know the fitness effects of the trait and the relatedness to the recipients. In practice, pedigree relatedness usually suffices (because it leads to the genes in the genome pulling in the same direction), making experiments surprisingly feasible (West and Gardner, 2013). This is supported by the success of the vast body of empirical literature that has sprung from inclusive fitness theory (for an entry into that literature, see: Foster, 2009; Davies et al., 2012), and for an attempt to quantify such successes (Abbot et al., 2011, Tables 1 and 2).

2.5. Advantage 5: General applicability as to the empiricist

Hamilton (1964) made remarkably few assumptions (namely autosomal diploidy, outbreeding, semelparity, and weak selection). This means we can study populations in which there are many types of individuals with interactions occurring with any number of recipients. In general, our models, and therefore predictions, don't have to be custom fitted to each new species we study, especially useful considering we rarely know the genetic details to do so. This not only leads to more theoretically grounded empirical work, but provides for broad unification across the tree of life. An aim of any science is to have simple, overarching frameworks which work across specific details. In turn, this generality allows us to make and test comparative predictions, which hold across populations and species. Comparative statics are a bedrock of evolutionary biology, and the generality of Hamilton's theory lends itself to them (Darwin, 1871; Parker and Maynard Smith, 1990; Harvey et al., 1991; Harvey and Purvis, 1991; Hughes et al., 2008; Cornwallis et al., 2010; Davies et al., 2012; Fisher et al., 2017; Cornwallis et al., 2017). For a further discussion of extensions to Hamilton's original paper, which have attempted overcoming some of the few assumptions he made, see for example Queller (1992), Grafen (2006), and Gardner et al. (2011).

3. The challenge of non-additivity

So far we have focused on inclusive fitness under additivity. Things get more complicated when we allow for fitness effects to combine non-additively, a problem first dealt with formally in a general way by Queller (1985). Non-additivity can arise a number of ways. For example, consider a simple two player game, in which players can either cooperate, giving their partner a fitness increment b, or defect. In an additive game, individuals receive b from cooperators regardless

of their strategy. But in a non-additive, synergistic game, when two cooperators interact, they get an additional, (possibly negatively) synergistic effect, d (Queller, 1985). This might occur, for example, if two hunters are more (or less) than twice as good as one. Non-additivity arises when the fitness effects of social actions combine, either with the existing number of offspring of an individual, or with each other, in a non-linear manner. While effects on fecundity may often naturally be assumed to combine additively, effects on survival are more likely multiplicative.

3.1. Advantage 1: Predicting gene frequency change

The challenge non-additivity poses for inclusive fitness has been discussed since at least 1978 (Cavalli-Sforza and Feldman, 1978; Uyenoyama and Feldman, 1982; Karlin and Matessi, 1983; Queller, 1985). The problem is two-fold. First, where before we needed to only know fitness effects and relatednesses to predict gene frequency change, we now need to know genetic make-up of the population, including the frequency of the gene (which will change under selection). Second, it's no longer even clear how to define inclusive fitness. Take, for example, the two player game described above. Inclusive fitness requires isolating the effects of the focal individual's genotype. But what portion of the synergistic component, d, is the focal individual responsible for?

Without a good way to define inclusive fitness in these scenarios, many authors turn to naive versions of inclusive fitness, such as 'simple-weighted sum', which are definable under synergy. Simple-weighted sum sums an actor's whole fitness and the fitness of all other individuals, weighted by their relatedness, which leads to double-counting of fitness effects, and is, importantly, not inclusive fitness (Grafen, 1982). A number of authors have shown that, for example, in a simple non-additive two player game, such naïve versions of inclusive fitness wrongly predict the direction of gene frequency change (Grafen, 1979; Queller, 1985; Lehmann et al., 2015; Okasha and Martens, 2016b; Taylor, 2017).

Several authors (Grafen, 1979; Lehmann et al., 2015; Okasha and Martens, 2016b; Allen and Nowak, 2016; Frank, 2013) have also pointed out that using Hamilton's 'neighbour modulated fitness' resolves this problem in some scenarios (although these authors don't always acknowledge that they are dealing with NMF, instead referring to it by other names, such as 'Grafen-1979 payoff', which is just neighbour modulated fitness in a two player game). This is not surprising as neighbour modulated fitness is simply mean number of adult offspring (it adds to the focal individual's fitness the offspring it would receive if its social partners expressed the same phenotype, weighted by the probability that they will, i.e. the population frequency of altruism enhanced or diminished by relatedness). That it correctly predicts gene frequency change stems directly from the fact that offspring are how genes are passed on.

Neighbour modulated fitness's ability to make the right prediction under a wider range of circumstances has led several authors to suggest adopting offspring number (under its various guises) in place of inclusive fitness (Lehmann et al., 2015; Okasha and Martens, 2016b). However, we proceed to discuss the ways in which mean offspring number is inferior to inclusive fitness with regards to the other advantages.

3.2. Advantage 2: A design principle for individuals

First, offspring number is not a design principle. Hamilton's (1964) starting point was neighbour modulated fitness because selection acts through offspring number. He developed inclusive fitness, despite it requiring more assumptions, because of its conceptual and practical advantages. In particular, inclusive fitness offers a design principle (advantage 2). It provides a link between gene frequency dynamics and design, because organisms can appear designed to maximise their inclusive fitness.

The same cannot be said for offspring number. As mentioned earlier, a design principle implies that organisms should appear to adjust their phenotypes to maximise a given quantity (Okasha and Martens, 2016b). An organism cannot adjust its neighbour modulated fitness, as her value of neighbour modulated fitness is outside her control. Neighbour modulated fitness is determined by the genotypes, or identities of a focal individual's partners. Adjusting it would require adjusting partners' genotypes. By analogy, offspring number is equivalent to 'being-part-of-a-group-of-four-ness' as a design principle (as contrasted with a simple propensity to join a group). Inclusive fitness, on the other hand, is under the control of the individual – an offspring simply has to adjust its own phenotype to alter its inclusive fitness (West and Gardner, 2013). Hamilton (1964) showed that, at equilibrium, organisms should appear to be choosing traits with regards to inclusive fitness, and that this results from gene frequency change. Although critics have doubted that IF is under the individual's control in general, they do accept the principle under additivity (Lehmann et al., 2015; Okasha and Martens, 2016b; Birch, 2017a,b). Thus, we lose the design principle if we use neighbour modulated fitness, which sacrifices most of the utility of inclusive fitness.

3.3. Advantage 3: Interpreting behaviour

Further, even if we were to stop using inclusive fitness for constructing models and designing experiments, its interpretive advantage (3) means that we would still use it to generate ideas, choose systems to study, and interpret social behaviour, provided the effective tradeoff is still roughly given by relatedness. Inclusive fitness tells us that organisms should tradeoff others' offspring against their own at a rate r, for relatedness. These leads us to identify systems with relatedness asymmetries, large opportunities for helping or harming, unusual

sex ratios, and extreme population structures as systems that would be fruitful for study. It also points us to traits that might disprove our theory. Traits that don't appear to abide by that valuation deserve further attention. With regard to such exceptions, we would like to know how far the tradeoff value is pushed away from r. We would also like to know whether different loci in the organism have their critical r pushed to the same extent or even in the same direction.

If the tradeoff value is not changed much, and changes inconsistently at different loci, then the complications will not alter the predictions of inclusive fitness very much. This is why it is important that no one who offers alternatives offers a useful interpretive principle, or explains how far the existing principle is really compromised.

3.4. Advantage 4: Empirical testability

Even if we stopped using inclusive fitness to construct models, we would struggle to continue our empirical work. The reason is straightforward: when you stop using inclusive fitness, you start needing to know genetics. Here's why. To test a prediction from inclusive fitness theory, we must observe which individuals act and calculate $rb-c$ for those actors. We might use information on who acts (and who doesn't) to estimate b and c, by subtracting average offspring number of non-actors from actors, and of non-recipients from recipients. More generally, we can regress the average adult offspring number on (i) the number of actions taken and (ii) the number of actions received. Thus even without knowing genotypes, we can apply inclusive fitness.

For neighbour modulated fitness the situation is more complicated, and we need to know genotypes. In the very simplest haploid two-allele model, higher average neighbour modulated fitness of an allele, H, compared to an alternative allele, N, tells us that H will be selected. But if strategies include rare deviant behaviour, and therefore the opportunity to act occurs only with some small probability $\delta << 1$, then only knowing NMF and who acts is not sufficient; instead, genotypes are needed. This is because an actor will rarely be a recipient, and so the actors do worse, even though the trait may be favoured by acting on relatives. On the basis of phenotypes, those relatives are counted in among the non-actors, raising the NMF of non-actors (see more detailed discussion of rare deviant behaviour in Section 4). We would need to know genotypes to add them to the actors, and to show that possessing the tendency to act was beneficial.

To recover a maximisation principle in the field, then, we need genotypes. Then we can obtain maximisation of NMF by averaging over the NMF of H allele-bearers and comparing that with N-allele bearers. At this point, one familiar with modelling may be confused, as models of NMF include relatednesses, and a mathematically equivalent NMF version of Hamilton's rule can be extracted. However, in the field, relatednesses are not needed for NMF – a simple count of mean offspring number already includes this information. However,

knowing who to count requires knowing genotypes. Specifically, to average over bearers of the allele in question, we need to know the genotypes of the individuals we study (usually impractical) and the genetics of the trait in question (which we rarely do in practice).

On the other hand, inclusive fitness offers the biologist a measure of phenotype that predicts evolutionary change. Fortunately, the phenotypic gambit, of assuming we do not need to know the genetic architecture of a trait, has proved remarkably successful (Grafen, 1984; West and Gardner, 2013; Davies et al., 2012).

It is also worth noting that the NMF approach does not involve identifying the fitness consequences of a social action. Rather, we need to know only the genotype and the number of offspring. Indeed, one would conclude that a gene was spreading or not, and not know whether the cause was social behaviour or pathogen resistance or liver-enzyme activity. To study social behaviour, we *should* investigate how the actions of one individual affect the number of offspring of another – that is what social behaviour is. Alternatives to inclusive fitness, such as neighbour modulated fitness, don't offer the empirical utility of inclusive fitness.

3.5. Advantage 5: General applicability as to the empiricist

Inclusive fitness offers practical advantages to the modeller. Some authors (Nowak et al., 2010; Allen et al., 2013; Allen, 2015; Allen and Nowak, 2015; Nowak and Allen, 2015; Akçay and Van Cleve, 2016; van Veelen et al., 2017) have suggested abandoning inclusive fitness for what they refer to as 'standard natural selection' models, which track gene frequencies. This approach is good at predicting gene frequency change in mathematical models. However, it requires generating a custom model for each new biological scenario. Inclusive fitness is a single framework that works across systems, independent of many (though of course not all) details. Hamilton's original model is surprisingly general, allowing both the theoretician and the empiricist to apply the ideas to systems with arbitrary numbers of interactions and many different kinds of individuals. This degree of generality and unity is a rare and sought after gift in the sciences.

Of course, there will always be limitations to validity, and the more these are understood the better. Recent critiques of inclusive fitness (e.g. Nowak et al., 2010; Allen and Nowak, 2015) might possibly be put to good use in that direction, though few new issues of significance have been brought to light since 1978 (Cavalli-Sforza and Feldman, 1978).

4. Conditionality

Before we proceed to discussing practicalities for behavioural ecologists, a simple model will help illustrate some of the above points. It is often pointed out

that neighbour modulated fitness and inclusive fitness calculations are mathematically equivalent, but what is less often clearly articulated is how they become distinct in practice. Here, a model makes the distinction clear, and shows how conditional behaviour brings to light the difficulties of applying NMF.

All behaviour is conditional, and models incorporating conditionality are important for understanding one of the advantages of inclusive fitness. In the unconditional case usually studied of inclusive fitness in a grouped population, a standard infinite haploid model with groups of size n is first introduced, with p as the frequency of an altruism allele, A, and using r for relatedness in the simplest way we write the average number of other altruists in the group of a randomly chosen altruist as $n_A = (n-1)(r+(1-r)p)$, and the average number of altruists in the group of a randomly chosen non-altruist (B) as $n_B = (n-1)(1-r)p$. We assume an altruist suffers a cost of c and gives b to each other group member.

Hamilton (1964) identified two measures of fitness for predicting gene frequency change. Neighbour modulated fitness (NMF) is simply a measure of mean offspring number, which sums an individual's fitness in the absence of social interactions and the effects of all individuals in the population on that individual. Inclusive fitness (IF) sums baseline asocial fitness, and the effect the actor has on all individuals, including itself, weighted by relatedness. The mean IF and mean NMF of A and B in this model are

$$\text{NMF} = \begin{cases} 1 - c + n_A b & \text{altruist} \\ 1 + n_B b & \text{non-altruist} \end{cases}$$
$$\text{IF} = \begin{cases} 1 - c + (n-1)rb & \text{altruist} \\ 1 & \text{non-altruist} \end{cases}.$$

When we substitute as indicated for n_A and n_B we obtain

$$\text{NMF} = \begin{cases} 1 - c + (n-1)(r + (1-r)p)b & \text{altruist} \\ 1 + (n-1)(1-r)pb & \text{non-altruist} \end{cases}$$
$$\text{IF} = \begin{cases} 1 - c + (n-1)rb & \text{altruist} \\ 1 & \text{non-altruist} \end{cases},$$

and find that the mean differences for altruist minus non-altruist, which predict the spread of altruism, are

$$\text{DNMF} = -c + (n-1)rb$$
$$\text{DIF} = -c + (n-1)rb.$$

Thus, NMF and IF predict the spread of the altruism allele in exactly the same cases. However, note even in this simple case that NMF and IF differ: in

particular, NMF includes the altruism provided by the background fraction of altruists ($(n-1)(1-r)pb$) for both altruists and non-altruists. The sum of these terms is the diluting factor of Hamilton (1964), and its presence in a model is a sign that NMF rather than IF is used. For example, the important work of Rousset (2004) on the evolution of social behaviour in structured populations employs NMF. In recent work, in which altruists are always rare so that $p = 0$, the difference technically between NMF and IF can be hard to make (e.g. Lehmann et al., 2015).

However, in a conditional model, the difference remains very clear when altruists are rare. Now we amend our model so that in each group one individual is selected at random to be the potential altruist, and a random other individual is selected to be the potential recipient. The probability that the actor will be an altruist will be $p_A = 1/n + ((n-1)/n)(r + (1-r)p)$ for the group to which a randomly selected altruist belongs and $p_B = ((n-1)/n)(1-r)p$ for the group of a randomly selected non-altruist. We distinguish by suffices on fitnesses (NMF and IF) between the fitnesses of (i) potential altruists (suffix PA) (ii) potential recipients (PR) and (iii) unselected individuals (US). The NMF and IF are now

$$\begin{aligned}
\text{NMFUS} &= 1 \\
\text{IFUS} &= 1 \\
\text{NMFPR} &= \begin{cases} 1 + p_A b & \text{altruist} \\ 1 + p_B b & \text{non-altruist} \end{cases} \\
\text{IFPR} &= 1 \\
\text{NMFPA} &= \begin{cases} 1 - c & \text{altruist} \\ 1 & \text{non-altruist} \end{cases} \\
\text{IFPA} &= \begin{cases} 1 - c + rb & \text{altruist} \\ 1 & \text{non-altruist} \end{cases}.
\end{aligned}$$

Substituting as indicated for p_A and p_B we find that the mean differences for altruist minus non-altruist are

$$\begin{aligned}
&\text{DNMFUS} = 0 \quad \text{DNMFPR} = rb\left(\frac{n-1}{n}\right) + \frac{b}{n} \quad \text{DNMFPA} = -c \\
&\text{DIFUS} = 0 \quad \text{DIFPR} = 0 \quad \text{DIFPA} = -c + rb.
\end{aligned}$$

The obvious interpretations are that an altruist always reduces its NMF by its action, while IF predicts that an altruist will spread if $rb - c > 0$. One interesting question is which of these quantities is likely to be observable in the absence of genetic information, if all we can observe are the actions, and the offspring numbers of the individuals. We assume that by direct sequencing

or pedigree information or demographic modelling we can estimate r. By observing actual actors (AA) and actual recipients (AR), compared to uninvolved individuals, we can estimate the mean offspring number of US, AA and AR. This yields an estimate of b ($AR-US$) and c ($AA-US$), from which we can calculate inclusive fitnesses. However, owing to ignorance of p, we cannot estimate most of the NMF values.

A second point is that it is not true that selection favours altruism if the NMF of realised altruists is greater than the average NMF. We would need to average in the NMF of actual recipients, but we cannot distinguish the genetic altruists from the genetic non-altruists, so we do not know which recipients to include in that average. Thus the correct mathematical statement that NMF predicts gene frequency changes applies in the theoretical situation that we know the genotypes of all the individuals, but not in the common empirical situation where we can observe only the actions.

In a more realistic situation, in which altruism opportunities arise randomly across the groups, and in which the chance of taking up an opportunity is genetically multifactorial, the simplicity of the inclusive fitness approach remains, while the NMF approach becomes more and more enmired. The theoretical and usual empirical situations are thus very distinct, and these differences need to be respected.

This simple model illustrates that even if NMF works better in a wider range of theoretical scenarios, as has been pointed out for decades (Hamilton, 1964; Grafen, 1979; Lehmann et al., 2015) it may not be a useful practical tool. We now turn to the question of what behavioural ecologists can make of these challenges.

5. Practicalities for behavioural ecologists

The previous discussion suggests that, while offspring number is useful for predicting gene frequency change in mathematical models, for those interested in social behaviour and design, it is not a viable option. Offspring number, being outside the control of the individual, cannot be an individual level design principle. Further, measuring predictions using neighbour modulated fitness usually requires knowing the relationship between genotype and phenotype, and being able to measure genotypes, something that is for now impractical in the field (and usually the laboratory, too). How should whole organism biologists proceed, then, if they were to aim to work without using the concept of inclusive fitness? We see three options.

5.1. Abandon design

The limitations of inclusive fitness has led some authors to call for abandoning an individual level design principle altogether (Nowak et al., 2010; Doebeli,

2010; Allen et al., 2013; Allen, 2015; Allen and Nowak, 2015; Nowak and Allen, 2015; van Veelen et al., 2017). However, none of these authors provide (i) an alternative explanation for design, (ii) a consistent, unified way to generate predictions, or (iii) an adaptive principle that can be tested in the field and the laboratory. Instead, they offer no design principle, and suggest making custom models for each new situation, usually using metrics that will be impractical to measure empirically, such as genotypes and the relationship between genotype and phenotype. It is therefore unsurprising that inclusive fitness has a huge empirical literature and the alternatives essentially none (Krebs and Davies, 1978, 1987; Charnov, 1982; Krebs and Davies, 2009; West, 2009; Westneat and Fox, 2010; Davies et al., 2012). This strengthens Birch's resolution that inclusive fitness offers a useful organising framework, and goes further in highlighting its practical and empirical utility (Birch, 2017a,b).

While the alternative approaches are useful for theoretical models of gene frequency change, when it comes to social behaviour, we see exquisite design, which demands explanation. Further, theories must make predictions that can be tested on real organisms. To be clear, we mean that hypotheses about social behaviour are tested using the working hypothesis that inclusive fitness is maximised: we do not mean that it is usually possible to test whether inclusive fitness is in fact maximised. That would require the same kind of genetic information that we argue is currently vanishingly rare and likely to remain rare. For students of social behaviour, abandoning the design approach is not a viable option. Fortunately, the design approach has been spectacularly successful. A more detailed discussion of the utility of adaptationism can be found elsewhere (Welch, 2017; Gardner, 2017).

5.2. A non-additive Hamilton's rule

Another option is to rewrite Hamilton's rule so that it makes correct predictions. Hamilton's rule is an inclusive fitness tool used for predicting the direction of selection. As we have said, inclusive fitness is undefined to the extent that fitness effects are strictly non-additive. Some authors have pointed out that one option is to redefine components of Hamilton's rule to make it fully general, even allowing for non-additive interactions (Queller, 1985, 1992; Gardner et al., 2011; Queller, 2011; Rousset, 2015; Taylor, 2016; Lehmann et al., 2016; Taylor, 2017). In the standard approach, b and c are effects on offspring number, and r is a measure of genetic similarity between two individuals. If we replace these values with regressions on fitness, we recover a fully general Hamilton's rule, which does not require additivity and always correctly predicts the direction of evolutionary change (Queller, 1992; Gardner et al., 2011; Rousset, 2015). Depending on the causal breakdown we desire, non-additive effects can be incorporated into their own term (Queller, 1992, 2011), or, alternatively, we can leave the fitness effects (b and c) unchanged, but replace r with a higher

order relatedness coefficient, for example one that captures the relatedness between a focal individual and a pair of recipients (Taylor, 2016, 2017). Both of these are very valuable theoretical advances, showing the complete generality of Hamilton's rule when parameters are chosen correctly.

However, as various authors have pointed out (Birch and Okasha, 2014; Birch, 2014; Taylor, 2016, 2017; Allen and Nowak, 2016; Okasha and Martens, 2016a), the cost of this generality is a loss in simple interpretation of the terms. They can no longer be understood as simple effects on offspring number, we no longer have a simple interpretation of social behaviour, and, without knowing genetics, the parameters are no longer easily measurable in the field and in the laboratory. Recently some authors (Nowak et al., 2017) have confused this general, regression form of Hamilton's rule with the simple one discussed here. While it's true that this general form of Hamilton's rule (sometimes referred to as 'HRG', Birch, 2014) gains generality at the cost of empirical utility, the simple Hamilton's rule we've discussed (or versions of it), defined in terms of effects on offspring number, is the one that has been used to enormous empirical success, as outlined above and reviewed by, e.g., Foster (2009), Abbot et al. (2011) (Tables 1 and 2), Bourke (2011), and Davies et al. (2012). Indeed, critics of the general form of Hamilton's rule have not offered an alternative that rivals the empirical utility of standard Hamilton's rule. The regression approach is a powerful conceptual advance (Rousset, 2015), but not empirically useful in the usual situation that the genetics of the individuals studied are unknown.

5.3. Using additive inclusive fitness as an approximation

A final option, then, is to use additive inclusive fitness as an approximation, and remain alert to when this approximation will fail (and by how much). Grafen (1985) has a list of reasons why additivity is probably non-problematic in practice. We discuss one example here, explaining how mutations of rare but possibly large effect (similar to the population genetic notion of 'penetrance') can resolve the usual problems that arise from non-additivity. This resolution has been discussed by a number of authors in specific cases, and here we argue for it being a potentially general solution (Queller, 1996; Grafen, 1979; Birch, 2017a).

Non-additivity creates a problem for inclusive fitness in that fitness effects (and therefore, changes in gene frequency) are no longer wholly attributable to a focal genotype. For example, consider a simple two player game with discrete strategies, where each player can choose to play either Cooperate (to give b at cost c) or Defect, and where when two cooperators interact they receive an added effect, d. A cooperator will have many occasions on which she encounters another cooperator, and how likely these occasions are depends on the degree of relatedness, or assortation, in the population (r). If we imagine a mutant in the population that played Cooperate instead of Defect, increasing r increases the

likelihood that its partner's strategy will also be Cooperate, and inclusive fitness fails to take this alteration in the partner's behaviour into account. As a result, a naïve version of inclusive fitness makes the wrong prediction in a discrete, non-additive, two-player game (Grafen, 1979; Okasha and Martens, 2016b).

However, if strategies are not discrete but continuous, where a player can choose to cooperate a fraction π of occasions, the situation changes. Now, a variant strategy plays Cooperate $\pi + \delta$ portion of occasions. In other words, it plays Cooperate instead of Defect on one occasion out of many, and the probability that it is the same occasion its related partner also plays Cooperate instead of Defect (because of the mutant strategy – it may often play Cooperate in absolute terms) is very low (Grafen, 1979).

This principle extends beyond simple two-player games. More generally, when the genetic component of the variability in how individuals act on any given occasion is proportionally low (which implies the δ-weak selection of Wild and Traulsen, 2007), we can use inclusive fitness to make accurate predictions. In this case, the only way r impacts the direction of selection is through an actor's vested interest in its social partners. When this type of variability is high, r also determines assortation of strategies, which inclusive fitness does not capture. Fortunately, a low genetic component of variability will be the norm for populations near equilibria, where it is usually reasonable to suppose we study organisms (Fisher, 1930; Grafen, 1985), a point endorsed by Birch 2017a; 2017b). Thus, for traits of interest to behavioural ecologists, inclusive fitness should often make the correct predictions even under non-additivity.

The mathematical importance of δ-weak selection has been discussed elsewhere (e.g. Wild and Traulsen (2007); Peña et al. (2015); Taylor and Frank (1996) Anonymous, *Submitted*). Our point here is to explain the kinds of biological scenarios that deliver this mathematical convenience, extending brief verbal arguments by Grafen (1979) and Queller (1996). In a companion paper (Anonymous, *Submitted*), we formalise this otherwise verbal argument and discuss two recent papers that look for inclusive fitness maximisation but fail to find it (Lehmann et al., 2015; Okasha and Martens, 2016b), both coming to the conclusion that expected offspring number ('u_B' in Lehmann et al. (2015) and 'Grafen 1979' in Okasha and Martens (2016b)) is a better measure. Anonymous, *Submitted* show that probabilistic mixing of phenotypes recovers inclusive fitness maximisation.

We also note that this type of probablistic mixing may also resolve some questions about how inclusive fitness moves from the level of the trait to the individual. Queller (1996) has argued that certain types of non-additivity can make defining inclusive fitness at the individual-level difficult, because different measures are required for different traits. Specifically, when individuals adopt different roles in an interaction, it is not always clear how to assign offspring

number to the control of one actor, analogous to the challenges of assigning offspring number when there is synergy between traits. In the absence of a formal analysis, we suspect that this type of non-additivity will also be resolved by allowing probabilistic mixing. In the meantime, we are reassured that Grafen (2006) allowed different types of social actions, including unique roles, and still recovered inclusive fitness maximisation.

Finally, we have assumed that intra-organismal conflicts (e.g. genomic imprinting) are not pulling organisms away from inclusive fitness optima. The effect of conflict on inclusive fitness equilibria is interesting, but beyond the scope of this paper (for an entry into that literature, see, e.g. Haig, 2002; Foster, 2011; Gardner and Úbeda, 2017). Genetic conflict would indeed very likely require genetic genetic knowledge to investigate.

5.4. Monitoring assumptions

Of course, effective non-additivity may not always hold. Fortunately, theory tells us what to be on the lookout for. For example, recent environmental change may mean populations are not near equilibria, and therefore additive genetic variability may be high. This is a caveat that applies to all evolutionary biologists, not just those studying social dilemmas. More specifically, we might suggest that students of social behaviour be on the lookout for clear assortation of actions in nature.

As we have said, non-additivity is problematic when there is strong assortation of actions, because inclusive fitness calculations don't take that additional effect of relatedness into account. This is something a field or laboratory worker can observe. For example, consider a population of birds in a wood. If relatives are not interacting, we wouldn't expect *strategies* to be correlated. However, if relatives do interact, but the genetic component of the variability in how individuals act on any given occasion is proportionally low (the δ-weak selection of Wild and Traulsen, 2007) we still wouldn't expect *actions* to correlate between interacting individuals. The reason, as stated above, is that the chances of two interacting relatives expressing the deviant action on the same occasion are low. To be clear, we use the phrase 'on the same occasion' to illustrate the point, but technically it does not refer to a set of different occasions that always arise (in which case non-additivity can arise even when only one individual possesses the trait), but rather to different possible occasions, only one of which arises (when non-additivity weakens because the chance of both being deviant is a lower order probability).

If we do observe clear assortation of deviant actions between partners in nature, it can be taken as a red flag that individuals may be engaged in a discrete game, and in this case, inclusive fitness may give the wrong answer (*if* the payoffs are also strongly non-additive). This kind of discreteness might be most likely to arise in bacteria, because they are more likely to have single

gene phenotypes. It may turn out that situations that generate problems for inclusive fitness are rare in nature. Either way, they do not require abandoning inclusive fitness. Instead, they serve as specific caveats for which to be on the lookout when conducting experiments.

It's worth considering one more aspect of the failure of inclusive fitness. Take, for example, situations in which inclusive fitness won't hold, due to high high additive genetic variability and strongly non-additive fitness effects. Are these exceptional cases consistent, in the sense that they make some consistent prediction as to how we should expect organisms to look or behave? Should they be more social than inclusive fitness predicts? Should they value the effects of their actions on others at $r+$ some predictable σ? Queller (1985) has suggested that in some cases, including simple two player games, the sign of the non-additive component, d, contains some information about the direction selection will proceed in.

More generally, two questions are relevant to empirical biologists exploring this issue. Is there some design principle other than inclusive fitness, or is inclusive fitness the central target, with exceptional cases unpredictably moving organisms off the mark in varying directions? And if there is some other central target, does it differ from inclusive fitness in a way we could reliably measure? We surmise, in the absence of relevant work, that deviations depend on details of the genetics in an unilluminating way (unless one happens to know the genetics), though of course we would very interested in any theoretical argument that claims to show the contrary.

6. Conclusion

If we are interested in exact predictions of gene frequency change in mathematical models, offspring number is the measure of fitness we should use. However, if we are interested in social behaviour and design, and in particular behaviour and design in nature, we should use inclusive fitness under approximate additivity. It does have some limitations. But the alternatives are worse. And despite its limitations, inclusive fitness has many great conceptual and practical advantages for biologists. Further, as we have argued here and illustrated elsewhere (Anonymous, *Submitted*), some of the theoretical limitations may disappear under biologically realistic scenarios.

If inclusive fitness is applicable, then all biological principles of social behaviour are equivalent to it. If inclusive fitness is not applicable, then we need to know genetics and therefore there can be no biological principle of social behaviour. Thus, the significant questions are: how good an approximation is the inclusive fitness approach, and does it allow the subject of social biology to exist? For the moment, it is consistent with what little we know that the

approximation is reasonable, and the empirical successes of social biology back up this conclusion.

Thus, the continuation of work with inclusive fitness is founded on a sophisticated notion of what assumptions are required for exactness of inclusive fitness, the consequences of likely deviations, and the assurance from empirical successes that the working hypothesis is by and large satisfactory. The cost of the nuance of this notion is that it is not *easily* captured in a fully general model. But it is conceptually more suited to the various roles inclusive fitness plays within biology than the mathematically general models of population geneticists. Not only is inclusive fitness a powerful organising framework (Birch, 2017a,b), but without it, we would have no useful theoretical approach for understanding social behaviour in the laboratory, in the field, and in comparative work.

7. References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., Andre, J.-B., Van Baalen, M., Balloux, F., Balshine, S., et al. (2011). Inclusive fitness theory and eusociality. *Nature*, 471(7339):E1.

Akçay, E. and Van Cleve, J. (2016). There is no fitness but fitness, and the lineage is its bearer. *Phil. Trans. R. Soc. B*, 371(1687):20150085.

Alexander, R. and Borgia, G. (1978). Group selection and the hierarchical organization of life. *Annual Review Ecological Systems*, 9:449–474.

Allen, B. (2015). Inclusive fitness theory becomes an end in itself.

Allen, B. and Nowak, M. A. (2015). Games among relatives revisited. *Journal of theoretical biology*, 378:103–116.

Allen, B. and Nowak, M. A. (2016). There is no inclusive fitness at the level of the individual. *Current Opinion in Behavioral Sciences*, 12:122–128.

Allen, B., Nowak, M. A., and Wilson, E. O. (2013). Limitations of inclusive fitness. *Proceedings of the National Academy of Sciences*, 110(50):20135–20139.

Anonymous (Sub.). Extending the range of additivity in using inclusive fitness.

Birch, J. (2014). Hamilton's rule and its discontents. *The British Journal for the Philosophy of Science*, 65(2):381–411.

Birch, J. (2017a). The inclusive fitness controversy: finding a way forward. *Royal Society open science*, 4(7):170335.

Birch, J. (2017b). *The philosophy of social evolution*. Oxford University Press.

Birch, J. and Okasha, S. (2014). Kin selection and its critics. *BioScience*, 65(1):22–32.

Bourke, A. F. (2011). The validity and value of inclusive fitness theory. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723):3313–3320.

Burt, A. and Trivers, R. (2006). Genes in conflict: the biology of selfish genetic elements. *Cambridge, MA: BelknapHarvard*.

Cavalli-Sforza, L. L. and Feldman, M. W. (1978). Darwinian selection and "altruism". *Theoretical population biology*, 14(2):268–280.

Charnov, E. L. (1982). *The theory of sex allocation*. Princeton University Press.

Cornwallis, C. K., Botero, C. A., Rubenstein, D. R., Downing, P. A., West, S. A., and Griffin, A. S. (2017). Cooperation facilitates the colonization of harsh environments. *Nature ecology & evolution*, 1(3):0057.

Cornwallis, C. K., West, S. A., Davis, K. E., and Griffin, A. S. (2010). Promiscuity and the evolutionary transition to complex societies. *Nature*, 466(7309):969.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Davies, N., Krebs, J., and West, S. (2012). *An introduction to behavioural ecology 4th edition*. Oxford: Wiley-Blackwell.

Doebeli, M. (2010). Inclusive fitness is just bookkeeping. *Nature*, 467(7316):661–661.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T., and West, S. A. (2017). The evolution of host-symbiont dependence. *Nature communications*, 8:15973.

Foster, K. (2009). A defense of sociobiology. In *Cold Spring Harbor symposia on quantitative biology*, volume 74, pages 403–418. Cold Spring Harbor Laboratory Press.

Foster, K. R. (2011). The sociobiology of molecular systems. *Nature Reviews Genetics*, 12(3):193.

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. *Journal of Evolutionary Biology*, 26(6):1151–1184.

Gardner, A. (2017). The purpose of adaptation. *Interface focus*, 7(5):20170005.

Gardner, A. and Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of evolutionary biology*, 22(4):659–671.

Gardner, A. and Úbeda, F. (2017). The meaning of intragenomic conflict. *Nature ecology & evolution*, 1(12):1807.

Gardner, A., West, S. A., and Wild, G. (2011). The genetical theory of kin selection. *Journal of evolutionary biology*, 24(5):1020–1043.

Grafen, A. (1979). The hawk-dove game played between relatives. *Animal behaviour*, 27:905–907.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, 298(5873):425.

Grafen, A. (1984). Natural selection, kin selection and group selection. *Behavioural ecology: An evolutionary approach*, 2:62–84.

Grafen, A. (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology*, 2(2).

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*, 238(3):541–563.

Haig, D. (2002). *Genomic imprinting and kinship*. Rutgers University Press.

Hamilton, W. D. (1964). The genetical theory of social behavior. i and ii. *Journal of Theoretical Biology*, 7(1):1–52.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, 228(5277):1218–1220.

Harvey, P. H., Pagel, M. D., et al. (1991). *The comparative method in evolutionary biology*, volume 239. Oxford university press Oxford.

Harvey, P. H. and Purvis, A. (1991). Comparative methods for explaining adaptations. *Nature*, 351(6328):619.

Hughes, W. O., Oldroyd, B. P., Beekman, M., and Ratnieks, F. L. (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science*, 320(5880):1213–1216.

Karlin, S. and Matessi, C. (1983). The eleventh ra fisher memorial lecture-kin selection and altruism. *Proceedings of the Royal society of London. Series B. Biological sciences*, 219(1216):327–353.

Kingman, J. F. (1961). On an inequality in partial averages. *The Quarterly Journal of Mathematics*, 12(1):78–80.

Krebs, J. R. and Davies, N. B. (1978). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Krebs, J. R. and Davies, N. B. (1987). An introduction to behavioral ecology. 2nd. *John Wiley & Sons*.

Krebs, J. R. and Davies, N. B. (2009). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Lehmann, L., Alger, I., and Weibull, J. (2015). Does evolution lead to maximizing behavior? *Evolution*, 69(7):1858–1873.

Lehmann, L., Mullon, C., Akcay, E., and Van Cleve, J. (2016). Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution*, 70(8):1689–1702.

Lehmann, L. and Rousset, F. (2014). Fitness, inclusive fitness, and optimization. *Biology & Philosophy*, 29(2):181–195.

Leigh, E. G. (1977). How does selection reconcile individual advantage with the good of the group? *Proceedings of the National Academy of Sciences*, 74(10):4542–4546.

Marshall, J. A. (2015). *Social evolution and inclusive fitness theory: an introduction*. Princeton University Press.

Marshall, J. A. (2016). What is inclusive fitness theory, and what is it for? *Current opinion in behavioral sciences*, 12:103–108.

Nowak, M. A. and Allen, B. (2015). Inclusive fitness theorizing invokes phenomena that are not relevant for the evolution of eusociality. *PLoS biology*, 13(4):e1002134.

Nowak, M. A., McAvoy, A., Allen, B., and Wilson, E. O. (2017). The general form of hamilton's rule makes no predictions and cannot be tested empirically. *Proceedings of the National Academy of Sciences*, pages 5665–5670.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310):1057–1062.

Okasha, S. and Martens, J. (2016a). The causal meaning of hamilton's rule. *Royal Society open science*, 3(3):160037.

Okasha, S. and Martens, J. (2016b). Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games. *Journal of evolutionary biology*, 29(3):473–482.

Parker, G. A. and Maynard Smith, J. (1990). Optimality theory in evolutionary biology. *Nature*, 348(6296):27.

Peña, J., Nöldeke, G., and Lehmann, L. (2015). Evolutionary dynamics of collective action in spatially structured populations. *Journal of Theoretical Biology*, 382:122–136.

Queller, D. C. (1985). Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature*, 318(6044):366–367.

Queller, D. C. (1992). A general model for kin selection. *Evolution*, 46(2):376–380.

Queller, D. C. (1996). The measurement and meaning of inclusive fitness. *Animal Behaviour*, 1(51):229–232.

Queller, D. C. (2011). Expanded social fitness and hamilton's rule for kin, kith, and kind. *Proceedings of the National Academy of Sciences*, 108(Supplement 2):10792–10799.

Queller, D. C. (2016). Kin selection and its discontents. *Philosophy of Science*, 83(5):861–872.

Rousset, F. (2004). *Genetic Structure and Selection in Subdivided Populations*. Princeton University Press.

Rousset, F. (2015). Regression, least squares, and the general version of inclusive fitness. *Evolution*, 69(11):2963–2970.

Strassmann, J. E. and Queller, D. C. (2010). The social organism: congresses, parties, and committees. *Evolution: International Journal of Organic Evolution*, 64(3):605–616.

Taylor, P. (2016). Hamilton's rule in finite populations with synergistic interactions. *Journal of theoretical biology*, 397:151–157.

Taylor, P. (2017). Inclusive fitness in finite populations – effects of heterogeneity and synergy. *Evolution*, 71(3):508–525.

Taylor, P. D. and Frank, S. A. (1996). How to make a kin selection model. *Journal of Theoretical Biology*, 180(1):27–37.

Uyenoyama, M. K. and Feldman, M. (1982). Population genetic theory of kin selection. ii. the multiplicative model. *The American Naturalist*, 120(5):614–627.

van Veelen, M., Allen, B., Hoffman, M., Simon, B., and Veller, C. (2017). Hamilton's rule. *Journal of theoretical biology*, 414:176–230.

Welch, J. J. (2017). What's wrong with evolutionary biology? *Biology & Philosophy*, 32(2):263–279.

West, S. (2009). *Sex allocation*. Princeton University Press.

West, S. A. and Gardner, A. (2013). Adaptation and inclusive fitness. *Current Biology*, 23(13):R577–R584.

West, S. A., Griffin, A. S., and Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology*, 20(2):415–432.

West, S. A., Griffin, A. S., and Gardner, A. (2008). Social semantics: how useful has group selection been? *Journal of Evolutionary Biology*, 21(1):374–385.

Westneat, D. and Fox, C. W. (2010). *Evolutionary behavioral ecology*. Oxford University Press.

Wild, G. and Traulsen, A. (2007). The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247(2):382–390.

6

Extending the range of additivity in using inclusive fitness

Extending the range of additivity in using inclusive fitness

Abstract

Inclusive fitness is a concept widely utilised by social biologists as the quantity organisms appear designed to maximise. However, inclusive fitness theory has long been criticised on the (uncontested) grounds that other quantities, such as offspring number, predict gene frequency changes accurately in a wider range of mathematical models. Here we articulate a set of modelling assumptions that extend the range of mathematical models under which inclusive fitness maximisation holds. We extend recent analyses that failed to find inclusive fitness maximisation in a formal model. We show (i) that previous authors have not used the correct inclusive fitness, (ii) how to capture inclusive fitness correctly, and (iii) that under the assumption of probabilistic mixing of phenotypes, inclusive fitness is indeed maximised in these models. We hope that in articulating these modelling assumptions and providing formal support for inclusive fitness maximisation, we help bridge a widening gap between empiricists and theoreticians, demonstrating to mathematicians why biologists are content to use inclusive fitness, and offering one way to utilise inclusive fitness in general models of social behaviour.

Keywords: inclusive fitness, biological design, fitness maximisation, δ-weak selection, social evolution, population genetics

1. Introduction

Inclusive fitness is an individual-level quantity identified by Hamilton (1964), which he showed, under some assumptions, to increase due to the action of natural selection. Hamilton pointed out that adult offspring number is affected not just by the actions of an individual but by those of the individuals it interacts with. He observed that measuring those effects involves averaging over possible distributions of genotypes, which in turn involves knowing gene frequencies in the population, neither of which are simple or readily available calculations (Hamilton, 1964). Accordingly, he turned to an alternative metric, 'inclusive fitness', which involves taking the perspective of the focal individual and its effects on others (as opposed to others' effects on it).

Hamilton (1964) provided a verbal definition for inclusive fitness as follows: the sum of an individual's adult number of offspring after it has been 'stripped of all components which can be considered as due to the individual's social environment', and a weighted sum of the 'quantities of harm and benefit which the individual himself causes' to the offspring numbers of others. The weightings are degrees of relatedness. Relatedness is a measure of genetic similarity between two individuals (r=1 for identical twins, r=0 for random population member, including possibility of self in finite populations). The exact definitions of the fitness effects and of relatedness differ in different formal treatments. For nearly 40 years, at least within behavioural and evolutionary ecology, most field and laboratory workers have treated inclusive fitness as the quantity that organisms appear designed to maximise, and tailored their studies and experiments accordingly (summarised in, e.g. Westneat and Fox, 2010; Davies et al., 2012).

However, since at least 1978 (Cavalli-Sforza and Feldman, 1978), the concept of inclusive fitness has been controversial, criticised most notably for assuming additivity of fitness effects. The type of additivity we discuss here refers to how the effects of different social actions combine to affect one individual's offspring number. Under non-additivity, these effects may combine in some non-linear way. The well known challenge for inclusive fitness is that under non-additivity changes in gene frequency are no longer wholly attributable to a focal genotype. Since at least 1979 authors have pointed out that, in such scenarios, mean offspring number does a better job at predicting gene frequency change (Grafen, 1979). Two recent papers (Lehmann et al., 2015; Okasha and Martens, 2016b) potentially lend support to both claims, looking for inclusive fitness maximisation, failing to find it, and identifying mean offspring number (in the guise of neighbour-modulated fitness, the calculation Hamilton termed 'unweildy') as a more successful alternative. Readers are referred to Submitted (Anon) for a detailed discussion of why mean offspring number is not a useful maximand in practice, and therefore why extending the range of inclusive fitness's application would be of interest. Here we focus on the latter task.

We reanalyse these two models (Lehmann et al., 2015; Okasha and Martens, 2016b), articulating a new set of modelling assumptions which lead to the recovery of inclusive fitness maximisation. First, we point out that in both cases, the inclusive fitness measure used by the authors is not the natural one that arises from Hamilton's verbal definition. Second, we introduce a way of understanding Hamilton's verbal definition and capturing it mathematically that has been used before but not fully articulated (Grafen, 1979). The precise mathematical definition of inclusive fitness depends on the specific settings. However, in defining it precisely in specific cases below we aim to help mathematical biologists find the precise definition in their own setting. Third, we articulate a set of modelling assumptions that allows a very wide application of the additivity

assumption, showing that in the models developed by Lehmann et al. (2015) and Okasha and Martens (2016b) inclusive fitness is indeed maximised.

In doing so, we are following a recent resolution offered by Birch (2017), who argues that the critics (e.g. Nowak et al., 2010; Allen and Nowak, 2016; van Veelen et al., 2017) are right to point to technical difficulties in establishing that inclusive fitness is well-defined or that natural selection leads to 'as-if maximisation', *in a fully general theoretical model*. However, Birch argues that, within certain assumptions, notably additivity of fitness effects, inclusive fitness is close enough to being 'right' to justify its use as organising framework for understanding social behaviour. We strengthen Birch's resolution by extending the range of scenarios in which inclusive fitness can be applied. The significance of articulating modelling assumptions lies in the process by which biological ideas become transferred into mathematics. If biologists fail to explain clearly enough what they are doing, then the machinery of mathematics is capable of yielding unbiological answers that are hard for biologists to interpret or respond to. In controversies caused by failure of communication, biologists can be grateful for the work of philosophers in acting as intermediaries (Okasha and Martens, 2016a; Birch, 2017).

2. Okasha and Martens (2016)

Okasha and Martens (2016b) analyse a version of the Hawk-Dove game played between relatives (they focus on the simpler cooperation game, but we keep the discussion general here as the conclusions hold for both). Their goal was to look with mathematical precision at the question of whether inclusive fitness is maximised at equilibrium. Our first point is that neither of the two fitness functions they define corresponds to Hamilton's inclusive fitness, and we show what the third function is below. Our second point is that, when we allow all probabilistic mixtures of Okasha and Martens' strategies also to be strategies, this third function is indeed maximised.

2.1. Do they consider inclusive fitness?

Okasha and Martens' (2016b) first utility function, which they refer to as inclusive fitness, is

$$U(i,j) = V(i,j) + rV(j,i), \tag{1}$$

where r is relatedness, and $V(i,j)$ is an individual's payoff when playing strategy i against a partner who plays j. It is immediately apparent that this is not inclusive fitness, but something more akin to simple weighted sum fitness. It measures the actor's whole payoff plus r times its partner's whole payoff, and

therefore does not partition offspring number by causation. They find that this utility function is not maximised, which is not surprising, as it is not inclusive fitness.

Their second utility function, which they call 'Grafen 1979', is expressed as follows:

$$U\left(i,j\right)=rV\left(i,i\right)+(1-r)V(j,i). \tag{2}$$

This payoff function, identified by Grafen (1979), is simply mean number of offspring, and, as expected, Okasha and Martens (2016b) find that that the strategy with the highest value increases in frequency, and establish links between evolutionary dynamics and as-if maximisation. Clearly, neither of these utility functions is inclusive fitness as defined by Hamilton (1964).

2.2. What is the correct expression for inclusive fitness?

In order to ask whether inclusive fitness is maximised, we must write a third utility function, which sums the effect on personal payoff of expressing the strategy and the relatedness weighted difference to partner's payoff as a result of actor expressing the strategy, according to Hamilton's 1964 definition. To do this, we write k as a default, 'non-social' strategy, and therefore can express inclusive fitness as:

$$U(i,j)=V(i,k)+r\left(V\left(j,i\right)-V\left(j,k\right)\right). \tag{3}$$

Providing a 'non-social' strategy (k) allows us to apply Hamilton's definition of inclusive fitness, and this overcomes a problem that may have prevented mathematical modellers in the past from using that definition. A key additional point is that in testing for invasion, we employ the incumbent as the non-social phenotype. This device articulates an approach used by Grafen (1979, p.907).

2.3. Is inclusive fitness maximised under probabilistic mixing?

The problem of non-additivity remains. Consider the simple two player cooperation game with discrete strategies, analysed above, where each player can choose to play either Cooperate or Defect. Relatedness, r, is the measure of genetic similarity between players discussed above. In a simple two player game like this, r is also measures assortation between strategies. If we imagine a mutant in the population that played Cooperate instead of Defect, increasing r increases the likelihood that its partner's strategy will also be Cooperate, and inclusive fitness fails to take this alteration in the partner's behaviour into account. When fitness effects depend on the partner's genotype, as in the case of non-additivity, this oversight matters.

177
178 However, a simple verbal argument made elsewhere (Submitted, Anon; Grafen,
179 1979, final paragraph on p.906) suggests that this problem may dissolve in some
180 biologically relevant scenarios (note that although Grafen, 1979, explained the
181 second term in his equation (10) incorrectly, the formulae are correct). For ex-
182 ample, consider the case where strategies are not discrete but continuous, where
183 a player can choose to cooperate a fraction π of occasions. Now, a variant strat-
184 egy plays Cooperate on a fraction $\pi + \delta$ of occasions, where $\delta << 1$. In other
185 words, it plays Cooperate instead of Defect on one occasion out of many, and the
186 probability that it is the same occasion its related partner also plays Cooperate
187 is very low (Grafen, 1979).
188 Grafen (1979, p.907) has already shown that, when we allow for probabilis-
189 tic mixing of strategies, inclusive fitness correctly predicts the direction of gene
190 frequency change in the simple game above, and this resolves the problem iden-
191 tified by Okasha and Martens (2016b). In the appendix, we provide a proof for
192 this simple cooperation game, recovering the links between as-if inclusive fitness
193 maximisation and gene frequency change.
194 In summary, Okasha and Martens' 2016b 'inclusive fitness' function is not
195 inclusive fitness. The natural expression for inclusive fitness arising from Hamil-
196 ton's 1964) definition is our equation (3). Under probabilistic mixing, this cor-
197 rect inclusive fitness is indeed maximised.
198 **3. Lehmann et al. (2015)**
199
200 The game analysed above is a simple two player game. In some ways, this is
201 very general, because it allows us to make few assumptions about life cycle or
202 population structure (namely that the chance of meeting an identical strategy
203 depends only on r and p, which are both independent). However, it is a restricted
204 sort of interaction, in which r is a parameter rather than an endogenous feature
205 of the model. We now turn to a recent rigorous population genetic analysis,
206 which is in some ways more general, and which makes similar claims to Okasha
207 and Martens (2016b).
208 Lehmann et al. (2015) consider an infinite island model of haploid individ-
209 uals on patches of a fixed size. An individual's fitness is a function of its own
210 strategy and the profile of strategies to be found amongst its neighbours, where,
211 for our purposes, strategy sets are confined to real numbers. There is assumed to
212 be no class structure and there is permutation invariance of the elements of the
213 profile of neighbour strategies (which rules out associating with relatives more
214 than associating with random members of the group). Individuals are asexual
215 and haploid, and offspring migrate with some positive probability. Generations
216 may be discrete or overlapping, but adults do not migrate. Otherwise, no as-
217 sumptions are made about fecundity, survival, or competition. This allows for
218 any type of games to be played on the patches, and any type of strategies to
219
220

be employed. Thus, despite the highly specific population structure, the model is otherwise quite general, granting Lehmann et al. their 'wide latitude'. Accordingly, any conclusions drawn from the model about the maximisation of inclusive fitness are of interest.

The approach is then as follows. Consider a mutant individual and the conditions that must hold for this mutant strategy to invade the population. In an infinite island model, any mutant will either ultimately go extinct or go to fixation in the population. Thus, we can identify the uninvadability condition for a strategy. The question then becomes, is there a utility function for which the following is true: the strategies that maximise the utility function are the same as the strategies that satisfy the uninvadability conditions. If so, it can be said that, at equilibrium, organisms appear as if trying to maximise said utility function (Lehmann et al., 2015; Okasha and Martens, 2016b).

3.1. Do they consider inclusive fitness?

Lehmann et al. (2015) define three such candidate utility functions, but our first point will be that none of them corresponds to Hamilton's verbal definition of inclusive fitness with our interpretive principle, a fourth function that we exhibit below. Our second point is that, once we allow probabilistic mixtures of Lehmann et al.'s strategies also to be strategies, this fourth function is indeed maximised at equilibrium.

First, Lehmann et al. (2015) identify the utility function u_A, which they refer to as inclusive fitness.

$$u_A\left(x_i, \mathbf{x}_{-i}, 1_x\right) = w\left(x_i, \mathbf{x}_{-i}, 1_x\right) + r\left(\bar{x}, \bar{x}\right) \sum_{j \neq i} w\left(x_j, \mathbf{x}_{-j}, 1_x\right). \tag{4}$$

$w\left(x_i, \mathbf{x}_{-i}, 1_x\right)$ is the offspring number of an focal individual, i, expressing strategy x_i in a patch where the strategies of the individuals other than i are represented as $\mathbf{x}_{-i}$, where, recalling that the mutant is at zero frequency, the distribution of the whole population (1_x) is assumed to be monomorphic for x and $r\left(\bar{x}, \bar{x}\right)$ is relatedness (with $\bar{x}$ being the average strategy in the population). It is immediately apparent then that u_A is not inclusive fitness, but instead a version of 'simple-weighted sum' fitness (Grafen (1982)). It measures an individual's personal offspring number plus a weighted sum of the offspring of all its social interactants, and therefore fails to isolate the actor's effects, as Hamilton (1964) intended.

Lehmann et al. (2015) then turn to a second utility function, u_B, which they

refer to as 'average personal fitness',

$$u_B\left(x_i, \mathbf{x}_{-i}, 1_x\right) = \sum_{k=1}^{N} \sum_{\tilde{\mathbf{x}}_{-\mathbf{i}} \in P_k(\mathbf{x}_{-i})} w\left(x_i, \tilde{\mathbf{x}}_{-i}, 1_x\right) q_k\left(\bar{x}, \bar{x}\right), \tag{5}$$

where P_k is the subset of hypothetical neighbour strategy profiles such that $k-1$ neighbours have a strategy identical to the focal individual, and q_k is the probability of that profile (Lehmann et al., 2015). u_B is a version of neighbour modulated fitness (i.e. simply mean offspring number), as it counts an individual's offspring number incorporating the effects of its social partners. Note, of course, that Okasha and Martens' 'Grafen-79' payoff (our equation 2) is the simple two player game version of Lehmann et al.'s u_B (5), as both are mean offspring number. Lehmann et al. consider a third function, u_C, which we don't reproduce here as it is simply neighbour modulated fitness under a special assumption about the link between offspring number and material payoffs.

Lehmann et al. find a much closer fit between the uninvadability conditions from the dynamic model and the first and second order conditions for 'as-if' maximisation by individuals for u_B than u_A (Lehmann et al., 2015). This is not surprising, as we expect this to hold for mean offspring number, and it parallels Okasha and Martens' finding about Grafen-1979. However, none of these functions is inclusive fitness as Hamilton (1964) outlined, and therefore their analysis cannot satisfactorily interrogate inclusive fitness maximisation.

3.2. What is the correct expression for inclusive fitness?

Instead, we require a fourth function, which we will call u_{IF}. In line with Hamilton (1964), to obtain u_{IF} we must sum three components: baseline asocial fitness, the difference to personal fitness as a result of the strategy, and relatedness weighted difference to social partners' fitnesses as a result of the strategy. Recalling our principle, from the previous example, of adopting the incumbent as the non-social strategy for inclusive fitness purposes, this can be written,

- Baseline asocial fitness – the average for an x-player, so

  $$w\left(x, \mathbf{x}^{N-1}, 1_x\right),$$

  where $\mathbf{x}^{N-1}$ indicates that all other group members play x,

- The difference to own personal fitness as a result of action, y:

  $$+\, w\left(y, \mathbf{x}_{-i}, 1_x\right) - w\left(x, \mathbf{x}_{-i}, 1_x\right),$$

  where the distribution of $\mathbf{x}_{-i}$ is taken from that for y-players.

- The difference to others' personal fitnesses as a result of action, weighted by relatedness:

$$+ r\left(y,x\right)\sum_{j\neq i}\left(w\left(x,\mathbf{x}_{-i-j}y,1_x\right)-w\left(x,\mathbf{x}_{-i-j}x,1_x\right)\right),$$

where, $\mathbf{x}_{-i-j}y$ indicates that all but two play x, and one other plays y, meaning that in $\mathbf{x}_{-i-j}i$, $\mathbf{x}_{-i-j}$ is the distribution of the $N-2$ remaining members of a patch, taken from the 'experience' of the recipient of a y-player, and the i (x or y) is the known strategy of the other individual. $r(y,x)$ is relatedness from the perspective of a y player in a population of resident x players. The assumption of permutation invariance (made by Lehmann et al. (2015)) ensures that there is only one kind of relatedness on the patch, namely to a random other patch-mate.

In total, then, we can express the inclusive fitness goal function as:

$$\begin{aligned} u_{IF}\left(y,x\right) =& w\left(x,\mathbf{x}^{N-1},1_x\right) \\ &+ w\left(y,\mathbf{x}_{-i},1_x\right)-w\left(x,\mathbf{x}_{-i},1_x\right) \\ &+ r\left(y,x\right)\sum_{j\neq i}\left(w\left(x,\mathbf{x}_{-i-j}y,1_x\right)-w\left(x,\mathbf{x}_{-i-j}x,1_x\right)\right) \end{aligned} \tag{6}$$

3.3. Is inclusive fitness maximised under probabilistic mixing?

In section 2.3, we offered a verbal argument for the biological importance of low additive genetic variability. Here, we formalise this argument by extending Lehmann et al.'s (2015) model to allow for such continuous variation in strategies, and analyse the first and second order conditions for evolutionary uninvadability and maximisation of inclusive fitness (u_{IF}). We do this by considering a mutant strategy $\tilde{y}$, in which a mutant displays the deviant behaviour with some small probability, ϵ, and otherwise, with probability $(1-\epsilon)$, behaves like a resident (x), which we can write in a natural notation as $\tilde{y}=(1-\epsilon)x+\epsilon y$.

We proceed to ask whether evolutionary uninvadability = utility maximisation, which is true if the first and second order conditions for uninvadability and utility maximisation are the same. Following Lehmann et al. (2015), we write the lineage fitness of the mutant as:

$$W\left(\tilde{y},x\right)=\sum_{k=1}^{N}\binom{N-1}{k-1}q_k\left(\tilde{y},x\right)w\left(\tilde{y},\tilde{\mathbf{y}}^{(k-1)},\mathbf{x}^{(N-k)},1_x\right) \tag{7}$$

Where W is the lineage fitness of the mutant, w is the personal fitness of the mutant expressing $\tilde{y}$ in a patch with $k-1$ other mutants and $N-k$ residents

displaying x, in a population otherwise monomorphic for x. q_k is the probability that the neighbour profile of the focal mutant will consist of $k-1$ other mutants. Lehmann et al. (2015) have shown that, for x to be uninvadable, it must be that $y = x$ is a local maximum of W.

In the appendix, we find the first order condition for uninvadability under our probabilistic mixing condition (the first partial derivative of W) to equal the first order condition for utility maximisation (the first partial derivative of u_{IF}), and the same for the second order conditions. Therefore as a result of gene frequency dynamics, at equilibrium organisms appear as if trying to maximise inclusive fitness. Due to the 'wide latitude' afforded by the approach, this result holds some generality for inclusive fitness maximisation.

In summary, Lehmann et al. (2015) did not analyse inclusive fitness as defined by inclusive fitness. We derive the natural expression above in equation (6). We show in the appendix that under probabilistic mixing, the correct inclusive fitness is indeed maximised.

We expect our result to hold for other recent analyses which have identified mean offspring number as a successful maximand (e.g. Allen and Nowak, 2015), if we adopt our newly articulated modelling approach of regarding the incumbent as Hamilton's "nonsocial" case, and of allowing all probabilistic mixtures of elements in the original strategy set. An interesting future step would be to try to extend our result to more general population structures. Our articulation of these additional conditions will, we hope, help mathematicians and biologists understand each other better in future.

4. Discussion

Inclusive fitness has formed the bedrock of a vast body of empirical literature (for an entry into that literature, see: Foster (2009); Davies et al. (2012), and for an attempt to quantify such successes Abbot et al. (2011), Tables 1 and 2). However, it has long been criticised for its assumptions, most notably additivity of fitness effects, and its failure in such scenarios to predict gene frequency change as well as mean offspring number (sometimes referred to as 'neighbour-modulated fitness'). Recent papers have purportedly lent support to such claims with general mathematical models (Lehmann et al., 2015; Okasha and Martens, 2016b). However, we have shown that such models fail to correctly capture inclusive fitness, and that when the correct expression is used, under the assumption of probabilistic mixtures of phenotypes inclusive fitness maximisation is recovered.

The biological significance of the 'probabilistic mixtures' assumptions is important to understand. Some of what follows is at the moment our own intuition, and obtaining mathematical proofs of precise versions would be extremely useful. First, uncontroversially, it will usually be conceivable that the assumption

is true in any particular example, and cannot be ruled out. Second, we conjecture that the possible deviations from the assumption will not tilt the biology in any particular direction, and thus we can consider the equilibrium under the probabilistic mixtures assumption as a central case. The fact that this central case applies without knowledge of the genetics across such a wide range of possibilities is very important in regarding social biology as possible without detailed genetic knowledge. Finally, when the assumption is not true, we posit there will usually be no dynamic equilibrium: rather, alleles will be able to invade and change frequencies continually, with the behaviour of the population constantly changing. Furthermore, the behaviour predicted by the probabilistic mixtures model will be a focus of the orbits of the dynamically changing population mean behaviour.

We expect that the importance of probabilistic mixtures of phenotypes may extend to more general scenarios in which the genetic component of the variability in how individuals act on any given occasion is proportionally low (which implies the δ-weak selection of Wild and Traulsen, 2007), because it removes the assortation effect of r. We expect this scenario to be the norm for populations near equilibria (or, more precisely, near a point at which a monomorphic population is uninvadeable by any one of set of mutations that code for all nearby phenotypes), where it is usually reasonable to suppose we study organisms (Fisher, 1930; Grafen, 1985; Birch, 2017).

The technical mathematical requirements of building rigorous population genetic models are considerable. They often require focusing on quite detailed special cases that are in themselves quite complex, or on abstract mathematical concepts representing the limits of the proof, which are quite complex, and often require adopting a very precise mathematical mode of reasoning. It is not surprising that linking back to general concepts in less technically demanding areas of biology often seems to prove difficult. In extending these models and formalising our verbal arguments, we hope to make it easier for future modellers to make links to the general and verbally expressed conceptual theory when they build precise mathematical population genetic models.

Empirical successes provide some assurance that the working hypothesis of inclusive fitness is by and large satisfactory. Here we hope to have lent some formal support for such assurance. Further, we hope that our mathematical arguments will allow future modellers to have a better understanding of the reasons many biologists remain content to use inclusive fitness. This better understanding could even in principle allow productive interactions between subdisciplines of biology whose dialogue hitherto has been less productive than might be desired.

5. Acknowledgements

We thank Kevin Foster, Sir Charles Godfray, Ashleigh Griffin, Laurent Lehmann, Rafe Kennedy, Jennifer Perry, Stuart West, and two anonymous reviewers for valuable comments on the manuscript. SRL is funded by The Clarendon Fund, Hertford College, and NERC.

6. Author Contributions

All authors contributed equally to the work.

7. Data Accessibility

There are no data to be archived.

8. References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., Andre, J.-B., Van Baalen, M., Balloux, F., Balshine, S., et al. (2011). Inclusive fitness theory and eusociality. *Nature*, 471(7339):E1.

Allen, B. and Nowak, M. A. (2015). Games among relatives revisited. *Journal of theoretical biology*, 378:103–116.

Allen, B. and Nowak, M. A. (2016). There is no inclusive fitness at the level of the individual. *Current Opinion in Behavioral Sciences*, 12:122–128.

Arias, A., Gutierrez, E., and Pozo, E. (1990). Binomial theorem applications in matrix fractional powers calculation. *Periodica Polytechnica Transportation Engineering*, 18(1-2):75–79.

Birch, J. (2017). The inclusive fitness controversy: finding a way forward. *Royal Society open science*, 4(7):170335.

Cavalli-Sforza, L. L. and Feldman, M. W. (1978). Darwinian selection and "altruism". *Theoretical population biology*, 14(2):268–280.

Davies, N., Krebs, J., and West, S. (2012). *An introduction to behavioural ecology 4th edition*. Oxford: Wiley-Blackwell.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

Foster, K. (2009). A defense of sociobiology. In *Cold Spring Harbor symposia on quantitative biology*, volume 74, pages 403–418. Cold Spring Harbor Laboratory Press.

Gardner, A., West, S. A., and Wild, G. (2011). The genetical theory of kin selection. *Journal of evolutionary biology*, 24(5):1020–1043.

Grafen, A. (1979). The hawk-dove game played between relatives. *Animal behaviour*, 27:905–907.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, 298(5873):425.

Grafen, A. (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology*, 2(2).

Hamilton, W. D. (1964). The genetical theory of social behavior. i and ii. *Journal of Theoretical Biology*, 7(1):1–52.

Lehmann, L., Alger, I., and Weibull, J. (2015). Does evolution lead to maximizing behavior? *Evolution*, 69(7):1858–1873.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310):1057–1062.

Okasha, S. and Martens, J. (2016a). The causal meaning of hamilton's rule. *Royal Society open science*, 3(3):160037.

Okasha, S. and Martens, J. (2016b). Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games. *Journal of evolutionary biology*, 29(3):473–482.

Submitted (Anon).

van Veelen, M., Allen, B., Hoffman, M., Simon, B., and Veller, C. (2017). Hamilton's rule. *Journal of theoretical biology*, 414:176–230.

Westneat, D. and Fox, C. W. (2010). *Evolutionary behavioral ecology*. Oxford University Press.

Wild, G. and Traulsen, A. (2007). The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology*, 247(2):382–390.

Appendix A. Okasha and Martens

Okasha and Martens (2016b) analyse the simple cooperation game described in the text, in which an altruist donates b to its partner at a cost c, and when two cooperators are paired they each receive an additional benefit, d. Under our probabilistic mixing assumption, a mutant altruist (A) cooperates with

some small probability, ϵ, and otherwise, with probability $(1-\epsilon)$, behaves like a resident (S) and defects.

From the payoffs of the game outlined above, we can use the least squares regression approach (outlined by Gardner et al. (2011)) to extract parameters from the population data to determine the the change in frequency of an altruism allele (Δp). Eliminating higher orders of ϵ, the condition for altruism to increase in frequency is

$$rb - c > 0 \tag{A.1}$$

where r is the standard kin selection coefficient of relatedness (and the scaling factor, ϵ, has been eliminated because it does effect equilibria). If $rb - c < 0$ altruism decreases in frequency, and if $rb - c = 0$ there is no change in p.

From the payoffs of the game and our definition of inclusive fitness (equation 3), the Nash Equilibrium of the game is (A, A) if $rb - c > 0$. If $rb - c < 0$ then (S, S) is a Nash Equilibrium, and if $rb - c = 0$ all pairs of strategies are Nash Equilibria. Thus there is a correspondence between the evolutionary dynamics and as-if maximisation behaviour. Specifically, the conditions for a gene to spread in frequency correspond to the conditions for a strategy with the same effects to be a Nash Equilibrium. Thus we can consider organisms, at equilibrium, to appear as though maximising their inclusive fitness (Okasha and Martens, 2016b).

Appendix B. Lehmann et al.

Following Lehmann et al. (2015), and the assumptions outlined in the text, we consider an infinite island model of haploid individuals on patches of size N. We extend Lehmann et al.'s (2015) analysis, and consider a mutant strategy $\tilde{y}$, in which a mutant displays the deviant behaviour with some small probability, ϵ, and otherwise, with probability $(1-\epsilon)$, behaves like a resident (x): $\tilde{y} = (1-\epsilon)x + \epsilon y$.

Appendix B.1. First order conditions

We can rewrite equation 7, by the definition of q and p from Lehmann et al. (2015), as,

$$W(\tilde{y}, x) = \sum_{k=1}^{N} p_k(\tilde{y}, x)\, w\left(\tilde{y}, \tilde{\mathbf{y}}^{(k-1)}\mathbf{x}^{(N-k)}, 1_x\right), \tag{B.1}$$

where p_k is the probability that a randomly drawn mutant has $k-1$ other lineage members in its patch. A strategy x is uninvadable if, given x, $\tilde{y} = x$ is a local maximum of $W(\tilde{y}, x)$.

Taking the first derivative with respect to deviations in y, we get:

$$\left.\frac{\partial W\left(\tilde{y},x\right)}{\partial y}\right|_{y=x}=\sum_{k=1}^{N}\frac{\partial}{\partial y}p_k\left(\tilde{y},x\right)w\left(\tilde{y},\tilde{\mathbf{y}}^{(k-1)}\mathbf{x}^{(N-k)},1_x\right)\Big|_{y=x} \tag{B.2}$$

$$+\sum_{k=1}^{N}p_k\left(\tilde{y},x\right)\frac{\partial}{\partial y}w\left(\tilde{y},\tilde{\mathbf{y}}^{(k-1)}\mathbf{x}^{(N-k)},1_x\right)\Big|_{y=x} \tag{B.3}$$

The first term is equal to 0 because at $y=x$ we can factor out the fitness term, and $\sum_{k=1}^{N}\frac{\partial}{\partial y}p_k\left(\tilde{y},x\right)=\partial\left(1\right)=0$.

Turning to the second term, we now allow for a total k mutants and an independent chance, ϵ, of each of them displaying the deviant behaviour. This gives:

$$\left.\frac{\partial W\left(\tilde{y},x\right)}{\partial y}\right|_{y=x}= \tag{B.4}$$

$$\sum_{k=1}^{N}p_k\left(\tilde{y},x\right)\frac{\partial}{\partial y}\sum_{h=0}^{k}\binom{k}{h}\epsilon^h\left(1-\epsilon\right)^{k-h}\left(\frac{h}{k}w\left(y,x^{n-h}y^{h-1},1_x\right)+\frac{k-h}{k}w\left(x,x^{n-h-1}y^h,1_x\right)\right)\Big|_{y=x},$$

where h is the number of mutants displaying the deviant behaviour. We can express the binomial as,

$$\binom{k}{h}\epsilon^h(1-\epsilon)^{k-h}\approx\begin{cases}(1-k\epsilon+O(\epsilon^2)) & h=0\\ k\epsilon+O(\epsilon^2) & h=1\\ O(\epsilon^2) & h\geq 2\end{cases} \tag{B.5}$$

Eliminating higher order terms of ϵ gives:

$$\left.\frac{\partial W\left(\tilde{y},x\right)}{\partial y}\right|_{y=x}= \tag{B.6}$$

$$\sum_{k=1}^{N}p_k\left(\tilde{y},x\right)\frac{\partial}{\partial y}\left[\left(1-k\epsilon\right)w\left(x,x^{N-1},1_x\right)+k\epsilon\left(\frac{1}{k}w(y,x^{N-1},1_x)+\frac{k-1}{k}w(x,x^{N-2}y,1_x)\right)\right]\Big|_{y=x}$$

$$+O\left(\epsilon^2\right)$$

We now adopt the notational convention of Lehmann et al (2015) that $w(y,x_{-1},1_x)$ should be regarded as having $N+1$ arguments for the purpose of differentiation. Thus we can take a partial derivative of up to the $N+1th$ argument (though only actually use up to N). This allows us to take the derivative

of one individual's offspring number with respect to the behaviour of other single members of the group. By permutation invariance, and following Lehmann et al. (2015) in denoting w_j as the derivative of w with respect to its jth argument we get:

$$\frac{\partial W\left(\tilde{y},x\right)}{\partial y}\bigg|_{y=x} = \epsilon\left(\sum_{k=1}^{N} p_k\left(\tilde{y},x\right)\left(k-1\right) w_N\left(x,x^{N-1}y,1_x\right)\right) + \epsilon\left(\sum_{k=1}^{N} p_k\left(\tilde{y},x\right) w_1\left(y,x^{N-2},1_x\right)\right)\bigg|_{y=x} \tag{B.7}$$

And from the definition of relatedness, r (Lehmann et al. (2015)), $r\left(\tilde{y},x\right) = \sum_{k=1}^{N} \frac{p_k(\tilde{y},x)(k-1)}{(N-1)}$ we obtain a first order condition of:

$$\begin{aligned}\frac{\partial W\left(\tilde{y},x\right)}{\partial y}\bigg|_{y=x} = & \\ \epsilon\left(w_1\left(x,x^{N-1},1_x\right) + (N-1)\sum_{k=1}^{N}\frac{p_k\left(x,x\right)\left(k-1\right)}{(N-1)} w_N\left(x,x^{N-1},1_x\right)\right) & \\ = \epsilon\left[w_1\left(x,x^{N-1},1_x\right) + (N-1)\, r\left(x,x\right) w_N\left(x,x^{N-1},1_x\right)\right] = 0. &\end{aligned} \tag{B.8}$$

Appendix B.2. Second order condition

The second order condition is given by the second derivative:

$$\begin{aligned}&\frac{\partial^2 W\left(\tilde{y},x\right)}{\partial y^2}\bigg|_{y=x} = \\ &\sum_{k=1}^{N}\frac{\partial^2}{\partial y^2} p_k\left(\tilde{y},x\right) w\left(\tilde{y},\mathbf{y}^{k-1}\mathbf{x}^{N-k},1_x\right)\bigg|_{y=x} \\ &+ 2\sum_{k=1}^{N}\frac{\partial}{\partial y} p_k\left(\tilde{y},x\right)\frac{\partial}{\partial y} w\left(\tilde{y},\mathbf{y}^{k-1}\mathbf{x}^{N-k},1_x\right)\bigg|_{y=x} \\ &+ \sum_{k=1}^{N} p_k\left(\tilde{y},x\right)\frac{\partial^2}{\partial y^2} w\left(\tilde{y},\mathbf{y}^{k-1}\mathbf{x}^{N-k},1_x\right)\bigg|_{y=x}\end{aligned} \tag{B.9}$$

The first term of the RHS of the equation equals 0 because at $y = x$ we can factor out the fitness term, and $\sum_{k=1}^{N}\frac{\partial^2}{\partial y^2} p_k\left(\tilde{y},x\right) = \partial^2\left(1\right) = 0$.

Turning to the second term, we already have the partial derivative of w (above) as: $\epsilon k w_j\left(y^1,x^{N-1},1_x\right)$. Using the definition of p_k from Lehmann et al. (2015), we write:

$$\frac{\partial}{\partial y}p_k\left(\tilde{y},x\right)=\frac{\partial}{\partial y}\frac{kt_k\left(\tilde{y},x\right)}{\sum_{h=1}^{N}ht_h\left(\tilde{y},x\right)} \tag{B.10}$$
$$=\frac{\frac{\partial}{\partial y}kt_k\left(\tilde{y},x\right)\left(\sum_{h=1}^{N}ht_h\left(\tilde{y},x\right)\right)-kt_k\left(\tilde{y},x\right)\frac{\partial}{\partial y}\sum_{h=1}^{N}ht_h\left(\tilde{y},x\right)}{\left(\sum_{h=1}^{N}ht_h\left(\tilde{y},x\right)\right)^2},$$

where $t_k(\tilde{y},x)$ is the number of demographic periods ('sojourn time') the lineage consists of k individuals. To get an expression for $\frac{\partial}{\partial y}t_k\left(\tilde{y},x\right)$, we use the matrix, Q, from which t_k is derived, as defined in the Supplementary Material of Lehmann et al. (2015)). Q is a matrix whose i, jth entry is the probability a patch with j mutants becomes a patch with i mutants in the next demographic period. To obtain the formula, we need a symbol $R_{i-f,j-h,h}(y,x)$ for the probability that the $j-h$ non-deviant mutants contribute $i-f$ individuals to the next time step. The probability of going from j to i mutants is then

$$\begin{aligned}Q_{ij}(\tilde{y},x)=&\pi_{j0}Q_{ij}(x,x)+\\&\sum_{h=1}^{j-1}\pi_{jh}\left(1-\sum_{k=1}^{N}Q_{k,h}\left(y,x\right)+R_{i(j-h)h}(y,x)\right)+\\&\sum_{h=1}^{j-1}\pi_{jh}\sum_{f=1}^{i-1}\left(Q_{f,h}(y,x)+R_{(i-f)(j-h)h}(y,x)\right)+\\&\sum_{h=1}^{j-1}\pi_{jh}\left(Q_{i,h}(y,x)+1-\sum_{k=1}^{N}R_{k(j-h)h}(y,x)\right)+\\&\pi_{jj}Q_{ij}(y,x)\end{aligned} \tag{B.11}$$

where h represents the number out of a total j mutants that display the behaviour y, and π_{jh} is the probability that a group with j mutants will have h individuals displaying the behaviour. The Q matrices capture individuals contributed by the deviant displaying mutants, and the R matrices capture mutant individuals contributed by mutants acting as residents.

Now we apply our assumptions that only a small fraction ϵ of mutants display, and that the chances of displaying are all independent. This gives us

$$\pi_{jh}=\binom{j}{h}\epsilon^h(1-\epsilon)^{j-h}\approx\begin{cases}(1-j\epsilon+O(\epsilon^2)) & h=0\\ j\epsilon+O(\epsilon^2) & h=1\\ O(\epsilon^2) & h\geq 2\end{cases} \tag{B.12}$$

Substituting and eliminating higher orders of ϵ, we get:

$$Q_{ij}(\tilde{y},x) = Q_{ij}(x,x) + j\epsilon\left(-Q_{ij}(x,x) + 1 - \sum_{f=i+1}^{N} Q_{f,1}(y,x) + 1 - \sum_{k=i+1}^{N} R_{k(j-1)1}(y,x)\right) \tag{B.13}$$

If **A** is the NxN matrix of the form:

$$\begin{bmatrix} 0 & 1 & 1 & ... & 1 \\ 0 & 0 & 1 & ... & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & ... & 1 \\ 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

and **B** is the NxN matrix of the form:

$$\begin{bmatrix} 1 & 1 & 1 & ... & 1 \\ 0 & 0 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

and **C** is the NxN matrix of the form:

$$\begin{bmatrix} 0 & 1 & 0 & ... & 0 \\ 0 & 0 & 1 & ... & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & ... & 1 \\ 0 & 0 & 0 & ... & 0 \end{bmatrix}$$

And writing **J** for the NxN matrix with the vector j as the diagonal and otherwise zeroes, and **1** as an NxN matrix of ones, and R_1 for the matrix with entries $R_{k,j,1}$ (representing the probability a patch with j non-deviant mutants, in the presence of 1 deviant mutant, becomes a patch with k mutants), we can write Q as:

$$Q(\tilde{y},x) = Q(x,x) + \epsilon\left(-Q(x,x) + \mathbf{1} - \mathbf{A}\left(Q\left(y,x\right)\mathbf{B}\right) + \mathbf{1} - \left(\mathbf{A}R_1(y,x)\right)\mathbf{C}\right)\mathbf{J} \tag{B.14}$$

Now, $t_k\left(\tilde{y},x\right)$ is taken from the first column of $\left(I - Q(\tilde{y},x)\right)^{-1}$. From the binomial theorem for matrices (Arias et al. (1990)),

$$(I - Q(\tilde{y}, x))^{-1} = (I - Q(x, x))^{-1} - \epsilon (I - Q(x, x))^{-1} ((Q(x, x) - \mathbf{\Omega}) \mathbf{J}) (I - Q(x, x))^{-1} + o(\epsilon^2), \tag{B.15}$$

where $\mathbf{\Omega} = \mathbf{1} + \mathbf{A} (Q (y, x) \mathbf{B}) - \mathbf{1} + (\mathbf{A} R_1(y, x)) \mathbf{C}$. Thus, writing $T(\tilde{y}, x)$ as the matrix $(I - Q(\tilde{y}, x)$ that contains as its first column the vector of t_k's, we write:

$$\frac{\partial}{\partial y} T (\tilde{y}, x) = -\epsilon \left(T (x, x) \frac{\partial}{\partial y} ((\mathbf{A} (Q (y, x) \mathbf{B}) + (\mathbf{A} R_1(y, x)) \mathbf{C}) \mathbf{J}) (T (x, x)) \right) \tag{B.16}$$

It follows from assumptions in Lehmann et al. (2015) (namely that x and y are real numbers) that the derivatives above are bounded except at a countable number of points, and thus the above expression is of order ϵ. Since the t_k's are taken from equation B.16, the derivative of t_k with respect to y is of order ϵ, which means that equation B.10 is of order ϵ, and thus line 3 of equal B.9 is of order ϵ^2. This gives:

$$\left. \frac{\partial^2 W (\tilde{y}, x)}{\partial y^2} \right|_{y=x} = \left. \sum_{k=1}^{N} p_k (\tilde{y}, x) \frac{\partial^2}{\partial y^2} w \left(\tilde{y}, \mathbf{y}^{k-1} \mathbf{x}^{N-k}, 1_x \right) \right|_{y=x} \tag{B.17}$$

Following the same steps as above (equations B.4-B.8) , we can write the second order condition as:

$$\begin{aligned} &\left. \frac{\partial^2 W (\tilde{y}, x)}{\partial y^2} \right|_{y=x} \\ &= \left. \sum_{k=1}^{N} p_k (\tilde{y}, x) \frac{\partial^2}{\partial y^2} \left[(1 - k\epsilon) w \left(x, x^{N-1}, 1_x \right) + k\epsilon \left(\frac{1}{k} w(y, x^{N-1}, 1_x) + \frac{k-1}{k} w(x, x^{N-2} y, 1_x) \right) \right] \right|_{y=x} \\ &= \epsilon \left[w_{11} \left(x, x^{N-1}, 1_x \right) + (N-1) r (x, x) w_{NN} \left(x, x^{N-1}, 1_x \right) \right] < 0. \end{aligned} \tag{B.18}$$

Appendix B.3. Utility maximisation

Turning to our utility function, u_{IF}, wewrite $u_{IF}(\tilde{y}, x)$ as:

$$\begin{aligned} u_{IF} (\tilde{y}, x) = &w \left(x, \mathbf{x}^{N-1}, 1_x \right) \\ &+ w (\tilde{y}, \mathbf{x}_{-i}, 1_x) - w (x, \mathbf{x}_{-i}, 1_x) \\ &+ r (\tilde{y}, x) \sum_{j \neq i} (w (x, \mathbf{x}_{-i-j} \tilde{y}, 1_x) - w (x, \mathbf{x}_{-i-j} x, 1_x)) \end{aligned} \tag{B.19}$$

Noting that when ϵ is small, $r(\tilde{y}, x) = r(x, x)$, we find the first and second order conditions to be:

$$\left.\frac{\partial u_{IF}(\tilde{y}, x)}{\partial y}\right|_{y=x} = \epsilon\left[w_1\left(x, \mathbf{x}^{N-1}, 1_x\right) + r(x, x)(N-1) w_N\left(x, \mathbf{x}^{N-1}, 1_x\right)\right] = 0 \tag{B.20}$$

$$\left.\frac{\partial^2 u_{IF}(\tilde{y}, x)}{\partial y^2}\right|_{y=x} = \epsilon\left[w_{11}\left(x, \mathbf{x}^{N-1}, 1_x\right) + r(x, x)(N-1) w_{NN}\left(x, \mathbf{x}^{N-1}, 1_x\right)\right] < 0 \tag{B.21}$$

The first and second order conditions for uninvadability and utility maximisation are identical.

7

Honest signalling and the double counting of inclusive fitness

Honest signalling and the double counting of inclusive fitness

Abstract

Inclusive fitness requires a careful accounting of all the fitness effects of a particular behaviour. Verbal arguments can potentially exaggerate the inclusive fitness consequences of behaviour, by including the fitness of relatives that was not caused by that behaviour, leading to error. We show how this this 'double counting' error can arise, with a recent example in the signalling literature. In particular, we examine the recent debate over whether parental divorce increases parent-offspring conflict, selecting for less honest signalling. We found that, when all the inclusive fitness consequences are accounted for, parental divorce increases conflict between siblings, in a way that can select for less honest signalling. This prediction is consistent with the empirical data. More generally, our results illustrate how verbal arguments can be misleading, emphasising the advantage of formal mathematical models.

Impact Summary

Evolutionary theory predicts that organisms should adopt traits that increase their inclusive fitness, a measure of fitness that includes contributions to relatives, and this idea underpins a wide range of empirical biology. But knowing how traits actually impact inclusive fitness requires a careful accounting of all the effects of a trait. A common mistake in calculating inclusive fitness is known as 'double counting', in which the effects of a trait are counted twice, and this has been shown to lead to incorrect predictions. We show how this problem arises and how it can be

avoided. We illustrate the point with a recent paper about begging in birds, which uses inclusive fitness to claim that divorce should have no impact on how honest baby birds are when begging for food. We show that this paper commits the double counting error, and develop mathematical models which avoid the error. The models predict that divorce, via its effects on inclusive fitness, should cause baby birds to be less honest when signalling their need. This occurs because divorce causes future siblings to be only half-siblings. We show that this prediction is supported by the empirical data. More generally, our results illustrate the risks of verbal reasoning and the benefits of developing mathematical models to clarify predictions. Inclusive fitness is a powerful tool in social biology which can guide empirical work and make testable predictions. This paper aims to help clarify how to use inclusive fitness correctly.

Introduction

Evolutionary theory predicts that selection for honest signalling can be reduced when there is greater conflict between individuals (Grafen, 1990; John Maynard Smith & Harper, 2003). This prediction can be hard to test with studies on single species, because the factors that determine conflict may not vary sufficiently to produce detectable variation (Popat et al., 2015). Caro et al. (2016) circumvented this problem with a comparative study across 60 species of birds, and examined whether greater conflict led to a weaker correlation between the intensity with which offspring beg and their long-term need (less honest signalling). In support of theory, Caro et al. (2016) found that offspring signalled less honestly when: (i) they face competition from current siblings; (ii) their parents are more likely to breed again; (iii)

parents are more likely to die or divorce (change mating partners between breeding bouts).

Bebbington & Kingma (2017) questioned one aspect of the third result with a verbal argument. Caro et al (2016) had argued that divorce should increase parent-offspring conflict because it means that future siblings, produced after the divorce, will only be half-siblings. This reduction in relatedness between siblings increases parent-offspring conflict (Hamilton 1964; Trivers 1974). In contrast, Bebbington & Kingma argued that divorce should have no impact on offspring honesty since an individual will gain two sets of half siblings, cancelling out the effects of losing one set of full siblings. Instead, they suggested a number of alternative hypotheses that could explain the data.

We show here that Bebbington & Kingma's argument makes what is termed the 'double-counting error' (Grafen, 1982, 1984; Queller, 1996). When considering the inclusive fitness consequences of divorce, they summed across all the siblings that were produced in the future. This leads to error because offspring are counted multiple times, as part of the fitness of multiple individuals. Instead, when considering the evolution of a trait, we need to focus on the specific consequences of variation in that trait, the 'inclusive fitness effect' (Frank, 1998; Grafen, 1982, 1984; Hamilton, 1964; Queller, 1996; Taylor, 1989, 1990; West, Griffin, & Gardner, 2007).

We present a simple inclusive fitness model to show how Bebbington & Kingma's conclusion was based upon double-counting. We then use a neighbour-modulated fitness approach to model this case more formally. Finally, we consider empirical support for the alternative hypotheses proposed by Bebbington &

Kingma. Our overall aim is to use an analysis of this particular problem to examine more general issues about how problems such as double counting can arise from simple verbal arguments.

Inclusive fitness and double counting

Caro et al. (2016) argued that if the parents of an individual divorce, then that individual will be half as related to future siblings, and so will be selected to obtain more resources in the short term from their parents, through less honest signalling. Bebbington & Kingma (2017) argued that this prediction should not hold, because divorce would also lead to that individual having twice as many siblings, because each parent goes on to raise a separate brood. Bebbington & Kingma argued that these two effects, twice as many offspring that are half as related, would exactly cancel, and so individuals should be indifferent to the likelihood that their parents will divorce.

However, what matters in the eyes of selection is the inclusive fitness *effect* of a trait, and not the total number of relatives produced (Hamilton 1964). Inclusive fitness does not include all the offspring produced by relatives, only those which are a result of the behaviour of the individual whose fitness we are measuring (see figure 3 in West et al. 2007, or Box 11.4 of Davies et al. 2012). So for example, if the helping behaviour of an actor leads to the beneficiary of that help producing another offspring, then that offspring would be counted in the inclusive fitness of the actor (indirect benefit) but not the beneficiary. To count that offspring both times, or even more if we also considered other relatives, is the double-counting error (Grafen

1982; Queller 1996). Bebbington & Kingma (2017)'s argument makes this double-counting error.

To illustrate this in simple terms, imagine a baby bird that signalled that it needed less food, and hence provided a marginal fitness benefit *B* to each parent, and the parents pass this benefit on future offspring (Figure 1). Put simply, the parent invests less in the current brood, and more in the future brood. In the case of monogamy, a baby is related to its future (full) siblings by 0.5, these receive a benefit of 2B (B from each parent), and therefore the total inclusive fitness effect is 2B x 0.5 = B. In the case of divorce, two sets of half-siblings, related by 0.25, each receive B, and the total inclusive fitness effect is 2 x 0.25 x B = 0.5B. Therefore, birds should be less honest in the case of divorce.

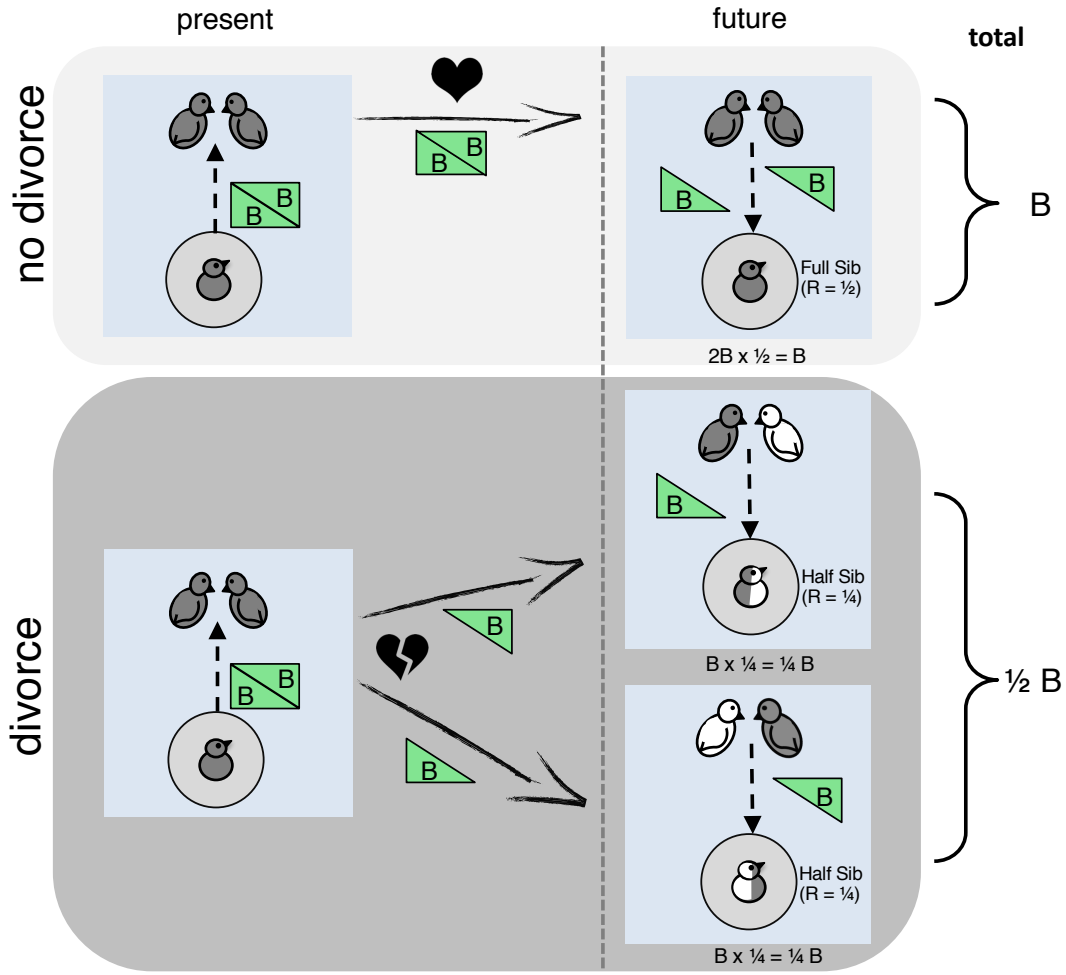

Figure 1. Divorce favours dishonesty because it causes offspring to be less invested in their future siblings. A proper counting of inclusive fitness requires isolating the direct effects of the trait, shown in green. Under monogamy, the effect the offspring has on its parents remains with both parents, and then is doled out to full siblings. In the case of divorce, the effects are divided between separated parents, and then doled out to half-siblings.

Bebbington & Kingma (2017)'s argument was wrong because it would require that the 2B given to the original parents translates to 2B in *each* of the remarriages. Phrasing the problem in terms of offspring number, without divorce a baby bird

gives, for example, an extra offspring to its mom and an extra offspring to its dad, which translates to two full siblings (*2 x 0.5*). Under divorce, this translates to two half-siblings (*2 x 0.25*). Bebbington & Kingma (2017)'s argument requires that, under divorce, the extra offspring given to mom and dad somehow double.

It is possible that that if divorce is common then the other parent may also be contributing an extra *B*, but that does not matter—it is not the consequence of the behaviour of the individual whose behaviour we are examining. Hamilton (1964) was the first to realise the potential for this confusion, and so he explicitly addressed it in his original definition of inclusive fitness, where he stressed the need to strip all components of fitness "*which can be considered as due to the individual's social environment*", and to focus on the "*fractions of the quantities of harm and benefit which the individual himself causes*". In the case we are considering, Hamilton's point means multiplying the benefit for future broods, after divorce, by *B* and not 2*B.* The potential for this double counting error has been highlighted by Grafen (1982; 1984), Queller (1996), and others (West et al. 207; Davies et al*.* 2012).

To provide another way of thinking about this problem, instead of a baby bird providing a benefit *B* to their parent, we can think of the baby bird as reducing the parents' overall resources. Each parent starts with V resources, and fitness is a function of the average resources of the parents. A baby can reduce its parents' resources by some fraction, f ($0<f<1$), such that they enter the next breeding season with (1-f)V resources. Assuming these resources are taken equally from each parent, a baby takes fV/2 resources from each parent. In the case of monogamy, a baby takes $fV/2 + fV/2 = fV$ from its full siblings ($r=0.5$), and therefore the effect is -

143 0.5fV. In the case of divorce, a baby take fV/2 from each of its two half sibling
144 broods (r=0.25), such that the total effect is -0.25fV. From an inclusive fitness
145 perspective, using up resources has a smaller negative inclusive fitness effect in the
146 case of divorce.
147
148 **A neighbour-modulated fitness model**
149 Inclusive fitness theory requires a careful accounting of all the fitness effects of a
150 particular behaviour, which can be complicated when reasoning verbally (Frank,
151 1998; Gardner, West, & Wild, 2011; P. D. Taylor, Wild, & Gardner, 2007; Peter D
152 Taylor & Frank, 1996). As illustrated above, and by previous discussions of the
153 double counting error, this could lead to behavioural consequences being
154 incorrectly added or missed (Grafen 1982; Queller 1996). A solution to this is to
155 develop theory with the neighbour-modulated fitness method of Taylor & Frank
156 (1996; Frank, 1997, 1998; Rousset, 2004; Taylor et al., 2007), which provides a
157 powerful and relatively simple way to derive an expression for the fitness
158 consequences of a behaviour.
159 We use the neighbour modulated fitness method to theoretically examine
160 whether the potential for divorce should influence the behaviour of an offspring. We
161 take a Maynard Smith (1991) approach, and deliberately develop a very simple
162 model, to illustrate the general point in an accessible way, rather than a more
163 complicated signalling model that would be less easy to follow.
164 We assume that there are only two years of breeding. There is a probability *d*
165 that parents 'divorce' between these two years, in which case they pair up with
166 another divorced parent in their second year. We assume that an offspring in the

first year of breeding can extract a proportion f of its parents' total resources, and that parents give the remaining $(1-f)$ of their resources to offspring in the second year. We wish to find if the amount of resources that the offspring should extract in their first year, *f*, is influenced by the divorce rate *d*.

The fitness, w, of an individual is a function of its own strategy (f), the strategy of its full sibling, F_{full}, the strategy of its half-sibling, F_{half}, and the population wide average, $\boldsymbol{F_{pop}}$ is:

$$w\left(f, F_{full}, F_{half}, \boldsymbol{F_{pop}}\right) = f\left((1-d)\left(1-F_{full}\right) + d\left(\frac{\left(\left(1-F_{half}\right)+\left(1-\boldsymbol{F_{pop}}\right)\right)}{2}\right)\right) \quad (1)$$

We wish to find the evolutionarily stable strategy (ESS), which is the strategy that cannot be beaten by any other strategy, and so would be stable under natural selection (Maynard Smith & Price, 1973). We assume that relatedness is equal to 1/2,1/4, and 0, for full siblings, half siblings and a random member of the population respectively. Using the Taylor and Frank's (1996) methodology, we find that the ESS is:

$$f^* = \frac{8}{3\left(4-d\right)} \quad (2)$$

In this case, divorce always matters, with increasing divorce rate causing babies to take more resources from their parents (Figure 2). This result formalises Caro et al.'s (2016) prediction that a greater likelihood of divorce leads to offspring being

favoured to extract more resources from their parents, and therefore being selected to signal less honestly.

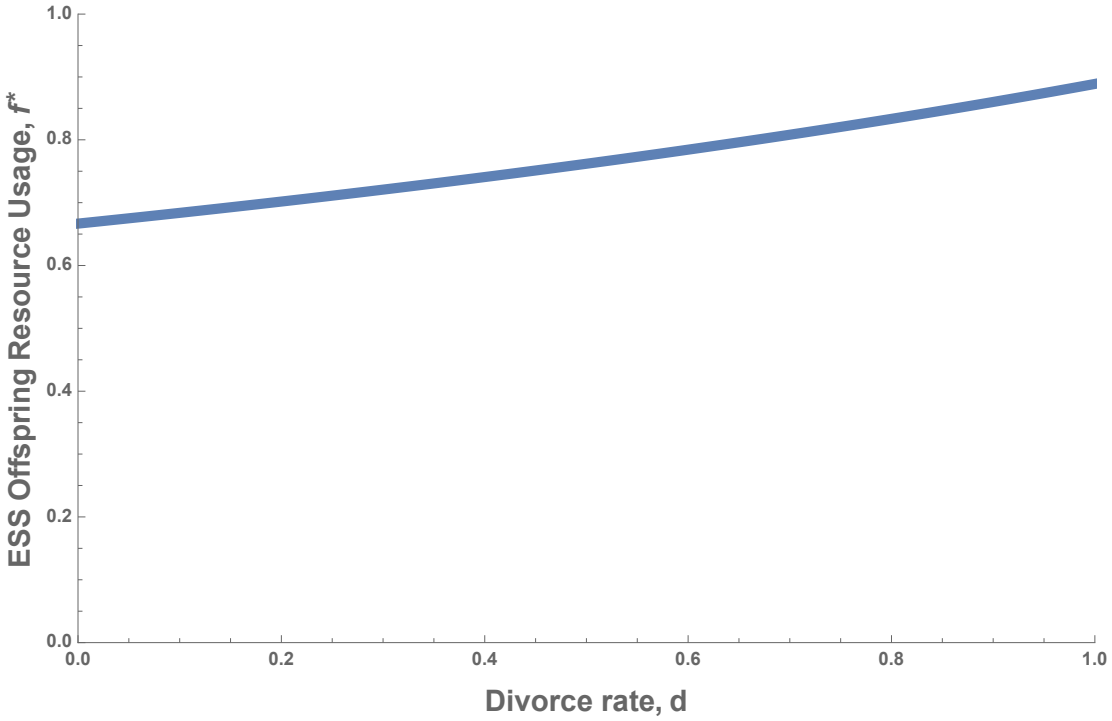

Figure 2. Divorce increases the optimal level of resources an offspring should take from its parents, a proxy for honesty of signalling (Equation 2).

The real world

Caro et al. (2016) found that offspring signalled less honestly when their parents were likely to divorce or die. They combined data from divorce and death because they shared a theoretical basis, with both leading to future offspring produced by parents being half-siblings. Furthermore, in that data set, there was no significant difference between the influence of divorce and death.

Based on the incorrect argument that divorce rates should not matter, Bebbington & Kingma (2017) proposed three alternative mechanisms that might be driving the patterns found by Caro et al. (2016):

(1) Pair bond duration could be confounded by clutch size and offspring competition. This is a valid concern, as brood size and the likelihood of parents breeding together again are correlated. However, Caro et al. (2016) specifically accounted for this by controlling for brood size in their analyses. Furthermore, we tested for collinearity by calculating variance inflation factors for Caro et al. 2016's model, and found low VIF values for all fixed effects well below the established cut-off of 10, or the more stringent cut-off of 3 (brood size VIF: 1.83; future reproduction VIF: 2.14; full vs. half siblings VIF: 2.44; Montgomery, Peck, & Vining, 2013; Zuur, Ieno, & Elphick, 2009). This indicates that brood size did not confound Caro et al. 2016's analyses.

(2) Divorce could be linked to competition for mates, making offspring dishonesty the result of higher levels of competitiveness in adults. There are no data to support this claim. Even if adult competitiveness was correlated with offspring competitiveness as a result of pleiotropy, selective pressures (e.g. divorce) acting on juvenile competitiveness should still shape behaviour, and we would still expect the qualitative differences in honesty predicted by Caro et al. (2016), More generally there is no theoretical reason to expect adults and offspring to be incapable of behaving differently at different times in their life. More generally, the suggestion that offspring and adult behaviour would have to be correlated in this way was an incorrect criticism of parent-offspring conflict theory (Dawkins, 1976; Godfray, 1995; Trivers, 1974).

(3) Divorce could be linked to parents' investing less in their current offspring, because short-term pair bonds raise conflict between parents ("scramble competition"). This is not an alternative to kin selection, as the influence it has on signalling is driven by an analogous decrease in relatedness. If conflict reduces the resources parents provide to offspring, this should enhance the effect of divorce on offspring dishonesty. Furthermore, even if this occurred, the effect would be small relative to the halving of relatedness cause by divorce.

The above discussion illustrates two general points regarding alternative explanations. First, we need to consider the effect size of alternative explanations. Given that divorce decreases the relative value of future siblings by ½, alternative explanations would need equally strong selective pressures to outweigh this influence. Second, hypotheses with more empirical support are more likely to be true. None of Bebbington & Kingma's (2017) alternative mechanisms have data to support them, whereas Caro et al.'s (2016) hypothesis does.

Future Extensions

Our above model was an idealised simplification, which aimed to illustrate the point that in the simplest case, divorce matters for offspring behaviour. There are a number of ways in which this model could be elaborated, to provide more specific predictions for scenarios of particular empirical interest. For example: the effects of signalling on parents' resources could be multiplicative, not additive; the effects on each parent might differ, potentially leading to intragenomic conflict; likelihood of divorce could vary with parental quality; or divorced parents might not breed again, or might breed with a lower quality individual. Another possibility is that death can

have a more complicated influence than divorce, because the death of one parent halves both relatedness to, and the number of future siblings.

Bebbington & Kingma (2017) considered one such extension, in which divorce raises the fitness of parents, following a comparative study by Culina et al. (2015). To eliminate the effect of divorce on offspring honesty, divorce would have to at least double the fitness of divorced parents (see our modelling section above). In contrast to this, Culina et al. found that divorce increased fitness by an average of only 37% more nestlings or fledglings, in a representative sample of 15 species. We investigated the possibility that the fitness consequences of divorce might eliminate the effect of divorce on honesty with an exploratory analysis on 15 species where there is data on both honesty and the fitness consequences of divorce. We found that, even when taking fitness consequences into account, and with a much smaller sample size, divorce still had a significant effect on offspring honesty (pMCMC = 0.0476*, n = 15 species, MCMCglmm model including phylogeny, study, species, brood size, future reproduction, the fitness consequence of divorce, and the likelihood of divorce and/or parental death). Future work could test the role of fitness consequences of divorce more thoroughly, or explore some of the other alternatives suggested above.

Conclusions

To conclude, we suggest two take home messages regarding the application of inclusive fitness theory to specific biological cases. First, care must be taken when formulating verbal predictions. Formal models can help resolve ambiguities and clarify predictions. Second, progress can be hindered when alternative mechanisms

or additional factors are mistaken for competing hypotheses. For example, the distinction between scramble competition and kin selection is a false one, as the former rests in part on the latter. Future progress is likely to be maximised by the interplay between theory and data.

Acknowledgements

SRL is supported by the Clarendon Fund, Hertford College, and NERC.

Author Contributions

SRL conceived of the manuscript and carried out the modelling. SMC carried out exploratory analyses. All authors contributed equally the preparation of the manuscript.

Data Accessibility

There are no data to be archived.

References

Bebbington, K., & Kingma, S. A. (2017). No evidence that kin selection increases the honesty of begging signals in birds. *Evolution Letters*. https://doi.org/10.1002/evl3.18

Caro, S. M., West, S. A., & Griffin, A. S. (2016). Sibling conflict and dishonest signaling in birds. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1606378113

299 Culina, A., Radersma, R., & Sheldon, B. C. (2015). Trading up: The fitness
300 consequences of divorce in monogamous birds. *Biological Reviews*, *90*(4),
301 1015–1034. https://doi.org/10.1111/brv.12143
302 Davies, N. B., Krebs, J. R., & West, S. A. (2012). *An introduction to behavioural*
303 *ecology 4th edition*. Oxford: Wiley-Blackwell.
304 Dawkins, R. (1976). *The Selfish Gene*. Oxford University Press.
305 Frank, S. A. (1997). Multivariate analysis of correlated selection and kin selection,
306 with an ESS maximization method. *Journal of Theoretical Biology*, *189*(3), 307–
307 316.
308 Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.
309 Gardner, A., West, S. A., & Wild, G. (2011). The genetical theory of kin selection.
310 *Journal of Evolutionary Biology*, *24*(5), 1020–1043.
311 Godfray, H. C. J. (1995). Evolutionary theory of parent–offspring conflict. *Nature*.
312 https://doi.org/10.1038/376133a0
313 Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, *298*(5873), 425.
314 Grafen, A. (1984). Natural selection, kin selection and group selection. *Behavioural*
315 *Ecology: An Evolutionary Approach*, *2*.
316 Grafen, A. (1990). Biological signals as handicaps. *Journal of Theoretical Biology*,
317 *144*(4), 517–546.
318 Hamilton, W. D. (1964). The genetical theory of social behavior. I and II. *Journal of*
319 *Theoretical Biology*, *7*(1), 1–52.
320 Maynard Smith, J. (1991). Honest signalling: the Philip Sidney game. *Animal*
321 *Behaviour*, *42*(6), 1034–1035. https://doi.org/10.1016/S0003-3472(05)80161-7
322 Popat, R., Pollitt, E. J. G., Harrison, F., Naghra, H., Hong, K. W., Chan, K. G., …

Diggle, S. P. (2015). Conflict of interest and signal interference lead to the breakdown of honest signaling. *Evolution*. https://doi.org/10.1111/evo.12751

Queller, D. C. (1996). The measurement and meaning of inclusive fitness. *Animal Behaviour*. https://doi.org/10.1006/anbe.1996.0020

Rousset, F. (2004). *Genetic structure and selection in subdivided populations*. Princeton University Press.

Smith, J. M., & Harper, D. (2003). *Animal signals*. Oxford University Press.

Smith, J. M., & Price, G. R. (1973). The Logic of Animal Conflict. *Nature*, *246*(5427), 15–18.

Taylor, P. D. (1989). Evolutionary stability in one-parameter models under weak selection. *Theoretical Population Biology*, *36*(2), 125–143. https://doi.org/10.1016/0040-5809(89)90025-7

Taylor, P. D. (1990). Allele-Frequency Change in a Class-Structured Population. *The American Naturalist*, *135*(1), 95–106.

Taylor, P. D., & Frank, S. A. (1996). How to make a kin selection model. *Journal of Theoretical Biology*, *180*(1), 27–37.

Taylor, P. D., Wild, G., & Gardner, A. (2007). Direct fitness or inclusive fitness: How shall we model kin selection? *Journal of Evolutionary Biology*, *20*(1), 301–309. https://doi.org/10.1111/j.1420-9101.2006.01196.x

Trivers, R. L. (1974). Parent-Offspring Conflict. *American Zoologist*, *14*(1), 249–264. https://doi.org/10.1093/icb/14.1.249

West, S. A., Griffin, A. S., & Gardner, A. (2007). Evolutionary explanations for cooperation. *Current Biology*, *17*(16), R661–R672.

8
Darwin's aliens

International Journal of Astrobiology

cambridge.org/ija

Darwin's aliens

Samuel R. Levin[1], Thomas W. Scott[1], Helen S. Cooper[2] and Stuart A. West[1]

[1]Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK and [2]37 Beech Croft Road, Oxford OX2 7AY, UK

Review

Cite this article: Levin SR, Scott TW, Cooper HS, West SA (2017). Darwin's aliens. *International Journal of Astrobiology* **0**, 1–9. https://doi.org/10.1017/S1473550417000362

Received: 30 May 2017
Accepted: 8 September 2017

Key words:
aliens; astrobiology; evolution; extraterrestrial life; individuality; major transitions

Author for correspondence:
Samuel R. Levin, E-mail: samuel.levin@zoo.ox.ac.uk

Abstract

Making predictions about aliens is not an easy task. Most previous work has focused on extrapolating from empirical observations and mechanistic understanding of physics, chemistry and biology. Another approach is to utilize theory to make predictions that are not tied to details of Earth. Here we show how evolutionary theory can be used to make predictions about aliens. We argue that aliens will undergo natural selection – something that should not be taken for granted but that rests on firm theoretical grounds. Given aliens undergo natural selection we can say something about their evolution. In particular, we can say something about how complexity will arise in space. Complexity has increased on the Earth as a result of a handful of events, known as the major transitions in individuality. Major transitions occur when groups of individuals come together to form a new higher level of the individual, such as when single-celled organisms evolved into multicellular organisms. Both theory and empirical data suggest that extreme conditions are required for major transitions to occur. We suggest that major transitions are likely to be the route to complexity on other planets, and that we should expect them to have been favoured by similarly restrictive conditions. Thus, we can make specific predictions about the biological makeup of complex aliens.

Introduction

There are at least 100 billion planets in our Galaxy alone (Cassan *et al.* 2012), and at least 20% of them are likely to fall in the habitable zone (Petigura *et al.* 2013), the region of space capable of producing a biosphere. Even if 0.001% of those planets evolved life, that would mean 200 000 life-harbouring planets in our Galaxy; and it would only take *one* alien life form for our conception of the Universe to change dramatically. It is no wonder, then, that hundreds of millions of dollars have recently been invested in astrobiology research (Schneider 2016), the USA and Europe have rapidly growing astrobiology initiatives (Des Marais *et al.* 2008; Horneck *et al.* 2016), and myriad new work has been done to try and predict what aliens will be like (Benner 2003; Davies *et al.* 2009; Rothschild 2009; Rothschild 2010; Shostak 2015). The challenge, however, is that when trying to predict the nature of aliens, we have only one sample – Earth – from which to extrapolate. As a result, making these predictions is hard.

So far, the main approach to making predictions about extra-terrestrial life has been relatively mechanistic (Domagal-Goldman *et al.* 2016). We have used observations about how things have happened on the Earth to make statistical statements about how likely they are to have happened elsewhere. For example, certain traits have evolved many times on the Earth, and so we posit that extraterrestrial life forms will converge on the same earthly mechanisms. Because eye-like organs have evolved at least 40 times (von Salvini-Plawen & Mayr 1977), and are relatively ubiquitous, we predict that they would evolve on other planets, too (Conway Morris 2003; Flores Martinez 2014). Similarly, we have used a mechanistic understanding of chemistry and physics to make predictions about what is most probable on other planets. For example, carbon is abundant in the Universe, chemically versatile, and found in the interstellar medium, so alien life forms are likely to be carbon-based (Cohen & Stewart 2001). These kinds of predictions come from a mixture of mechanistic understanding and extrapolating from what has happened on the Earth. There is no theoretical reason why aliens could not be silicon-based and eyeless.

An alternative approach is to use theory. When making predictions about life on other planets, a natural theory to use would be evolutionary theory. Evolutionary theory has been used to explain a wide range of features of life on the Earth, from behaviour to morphology. For example, it has allowed us to predict when some organisms, especially insects, should manipulate the sex of their offspring, to produce an excess of sons or daughters, how some birds should forage for food, and why males tend to be larger than females (Darwin 1871; Clutton-Brock & Harvey 1977; Davies & Houston 1981; West 2009; Davies *et al.* 2012). If life arises on other planets, then the evolutionary theory should be able to make similar predictions about it. Neither approach – theoretical or mechanistic – is more or less valid than the other. But each has different advantages and can be used to make different sorts of predictions.

CAMBRIDGE UNIVERSITY PRESS

Here, we examine how theoretical and mechanistic approaches can be combined to better understand what to expect from alien life. We consider whether aliens will undergo natural selection, and what implications would follow if they do. That aliens undergo natural selection is something often taken for granted, but which needs justification on firm theoretical grounds. We then turn our attention to a specific subset of aliens: complex ones. We examine how complexity has arisen on the Earth, and make predictions about how complexity would arise elsewhere in the Universe. Finally, we describe some biological features we would expect to find in complex extraterrestrial life.

Natural selection

On Earth

Darwin (1859) showed that just a few simple features of life on Earth lead to evolutionary change via natural selection. Individual organisms differ in how they look and act – there is natural *variation*. These differences are *heritable* – offspring tend to look and act like their parents. These heritable differences are linked to *differential success* – some individuals, as a result of how they are made or behave, leave more offspring than others. These three features, with heritable variation leading to differential success, result in natural selection (Darwin 1859; Fisher 1930). Any traits or behaviours linked to the greater production of offspring (higher fitness or success) will build up in the population over time. As the environment changes, different traits lead to higher success. This leads to changes in the population or evolutionary change.

Thus, the ingredients required for natural selection are incredibly simple. Given a collection of entities (a population) that has: (1) heredity; (2) variation; and (3) differential success linked to variation, then natural selection will follow. The entities that are more successful will become more prevalent in the population, as a result of being 'selected'. Natural selection does not depend on a specific genetic system (Darwin knew nothing of modern genetics) or a specific genetic material, elemental makeup or planet-type. Given that 1, 2 and 3 exist, natural selection occurs (Fig. 1).

Natural selection not only explains evolutionary change, it also explains adaptation. When we look around at the natural world, we cannot help but see what looks like design: a giraffe's neck is for reaching high up leaves, a stick insect's body for camouflage, a tree's leaf for photosynthesizing. Organisms look designed or 'adapted' for the world in which they live. Through the gradual selection of small improvements, traits associated with success in the environment accrue in the population. Consequently, over time, natural selection will lead to organisms that appear *as if* they were designed for success in the environment. The clause '*as if*' is key here – natural selection leads to the appearance of design (adaptation), without a designer (Grafen 2003; Gardner 2009).

In fact, natural selection is the *only* explanation we have for the appearance of design without a designer (Gardner 2009). Other processes can cause evolutionary change. For example, a mutation can cause a change from one generation to the next. But, without natural selection, random mutation is incredibly unlikely to produce the complex traits that we see around us, like limbs or eyes. Things that appear purposeful, such as limbs, organs and cells, require the gradual selection of improvements.

Another way to say this is that natural selection is unique because it is a *directional force*. The entities that increase in

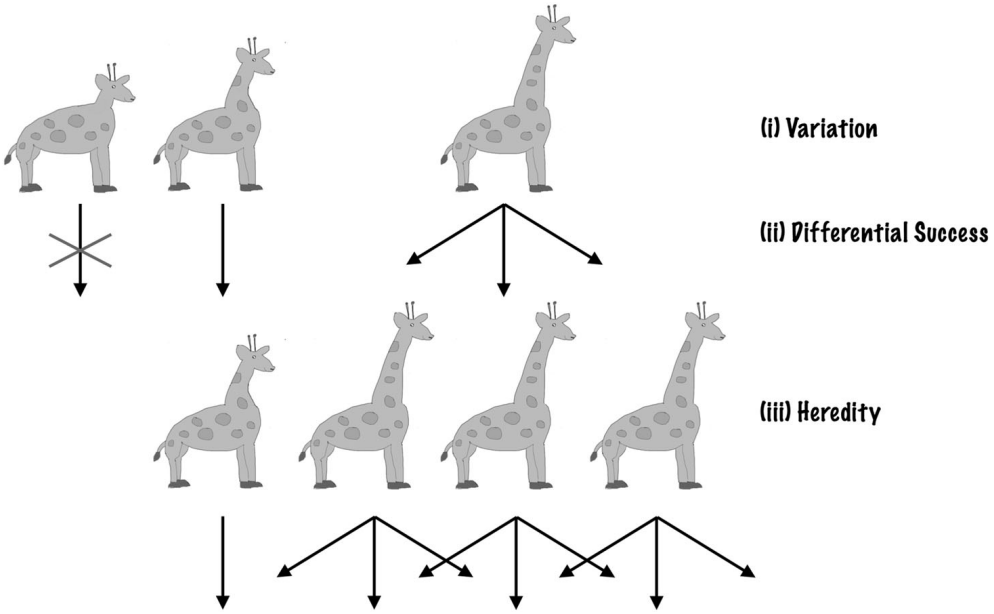

Fig. 1 – B/W online

Fig. 1. Natural selection. Natural Selection operates if three conditions are satisfied: variation, differential success linked to variation and heredity. Here, we illustrate with an example: the evolution of long necks in giraffes. (i) Initially, there are natural variations in giraffes' neck lengths. (ii) Longer-necked giraffes have access to more food, high up in the trees and so live longer to have more offspring. (iii) Giraffes' offspring resemble their parents. As a result of (i), (ii) and (iii), the population gradually shifts to be dominated by long-necked giraffes.

representation in the population are a *specific subset* of the population – those that are better at replicating. Natural selection increases fitness (Fisher 1930). As a result of these 'successful' entities accruing in the population, over time entities become adapted for the *apparent* purpose of success. They look like 'well-designed' machines, with the 'purpose' of their 'design' being successful replication.

In space

Natural selection is the only way we know to get the kinds of life forms we are familiar with, from viruses to trees. By familiar, we are not restricting ourselves to life forms that look earthly. Instead, they are familiarly life-like in the sense that they stand out from the background of rocks and gases because they appear to be busy trying to replicate themselves. A simple replicator could arise on another planet. But without natural selection, it won't acquire apparently purposeful traits like metabolism, movement or senses. It won't be able to adapt to its environment, and in the process, become a more complex, noticeable and interesting thing.

We can ask, then, will aliens undergo natural selection? Evolutionary theory tells us that, for all but the most transient and simple molecules, the answer is yes. Without a designer, the only way to get something with the apparent purpose of replicating itself (something like a cell or a virus), is through natural selection. Consequently, if we are able to notice it as life, then it will have undergone natural selection (or have been designed by something that itself underwent natural selection).

It is easy to quibble about the definition of life, and as some authors have pointed out, trying to do so can reveal more about human language than about the external world (Cleland & Chyba 2002). Our goal here is not to thoroughly define life. We adopt a functional stance – what separates life from non-life is its apparent purposiveness, leading to tasks such as replication and metabolism (Maynard Smith & Szathmáry 1995). Further, without natural selection, entities cannot adapt to their environment, and are therefore transient and will not be discovered. If we identified an extra-terrestrial entity that we deemed to be a foreign life form, but that had no degree of adaptedness, this prediction would not hold.

Picture an alien (Fig. 2). If what you are picturing is a simple replicating molecule, then this 'alien' *might* not undergo natural selection (Fig. 2a). For example, it could replicate itself perfectly every time, and thus there would be no variation, and it would never improve. Or it might have such a high error rate in replication that it quickly deteriorates. If we count things like that as life, then there could be aliens that do not undergo natural selection. But if you are picturing anything more complex or *purposeful* than a simple molecule, then the alien you are picturing has undergone natural selection (Fig. 2b). This is the kind of prediction that theory can make. Given heredity, variation and differential success, aliens will undergo natural selection. Or, more interestingly, without those three things, aliens could not be more complicated than a replicating molecule. Given an adapted alien, one with an appearance of design or purpose, it *will have undergone natural selection.*

Complexity

What is complexity?

We have established that aliens will undergo natural selection. It also seems reasonable that, given the sliding scale from replicating molecules to large creatures with many 'body parts', and beyond, some alien discoveries would be more interesting than others. In particular, the more complex the aliens we find, the more interesting and exciting they will be, irrespective of whether they appear anything like the life forms on the Earth. Something similar to a

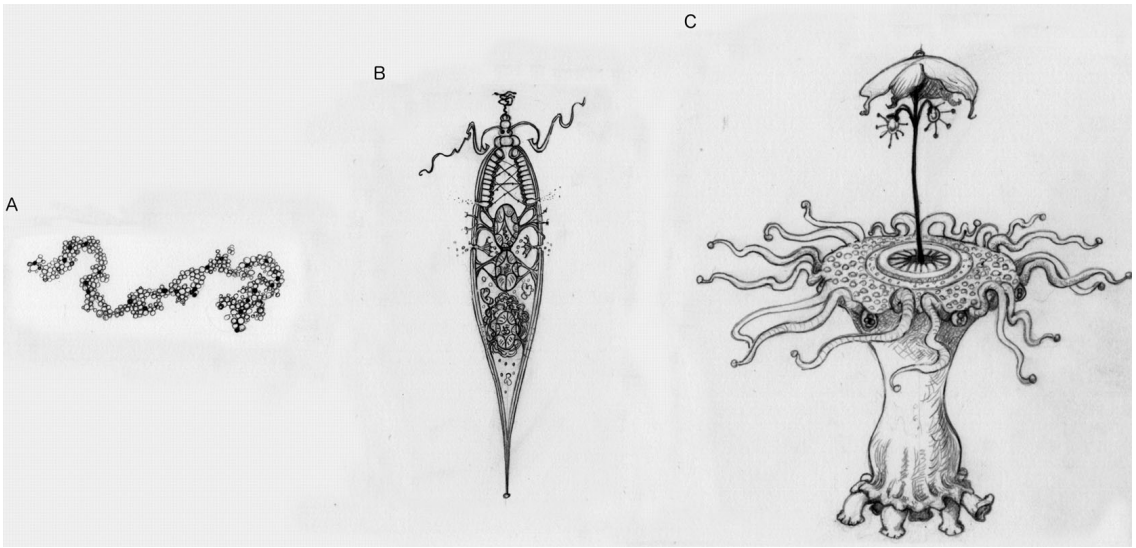

Fig. 2 - B/W online

Fig. 2. Picture an alien. These illustrations represent different levels of adaptive complexity we might imagine when thinking about aliens. (a) A simple replicating molecule, with no apparent design. This may or may not undergo natural selection. (b) An incredibly simple, cell-like entity. Even something this simple has sufficient contrivance of parts that it must undergo natural selection. (c) An alien with many intricate parts working together is likely to have undergone major transitions.

colony of Ewoks from Star Wars or the Octomite in Fig. 4 would likely be more interesting than a simple chemical replicator.

Complexity is difficult to define, and there is certainly no hard and fast rule about what is and is not complex. In biology, it is common to define complexity in terms of functional parts. Things with more parts taking on more tasks and containing more functional interactions are more complex (Maynard Smith & Szathmary 1995; Corning & Szathmáry 2015). A tree is more complex than a virus, and a beehive is more complex than a protein. Importantly, with organisms as with machines, the parts need to be working towards a common purpose, such as assembling a car or surviving to reproduce. Again, our goal here is not to provide definitions. The challenge comes at the boundaries, for example between a virus and a cell, where the definitions become murky. In the following sections, we are not focusing on the boundaries, but things, like the vast majority of life on the Earth, which clearly have a multitude of parts working in concert. Astrobiology is a largely empirical field, and the kinds of things programs like SETI are searching for are undeniably complex.

Complexity on Earth

What do we know about how complexity arises on the Earth? The theory of natural selection itself is silent about *whether* complexity will arise. The theory is useful for making predictions about what kinds of conditions or environments will lead to what kinds of evolutionary adaptations – not for making long-term predictions about the form of specific traits or creatures. However, recent advances in the field of evolutionary biology have shed light on how complexity has arisen on the Earth, on what points on the tree of life this has happened, and on what theoretical conditions favour it (Maynard Smith & Szathmáry 1995; Queller 1997; Bourke 2011; West *et al.* 2015).

In particular, the evolution of complex life on the Earth appears to have depended upon a small number of what have been termed major evolutionary transitions in individuality. In each transition, a group of individuals that could previously replicate independently cooperate to form a new, more complex life form or higher level organism. For example, genes cooperated to form genomes, different single-celled organisms formed the eukaryotic cell, cells cooperated to form multicellular organisms, and multicellular organisms formed eusocial societies (Maynard Smith & Szathmáry 1995; Queller 1997; Bourke 2011; West *et al.* 2015).

Major transitions

Major transitions on the Earth

Major evolutionary transitions are defined by two features. First, entities that were capable of replication before the transition can replicate only as part of a larger unit after it (interdependence). For example, the cells in our bodies cannot evolve back into single-celled organisms. Second, there is a relative lack of conflict within the larger unit, such that it can be thought of as an organism (individual) in its own right (Queller & Strassmann 2009; West *et al.* 2015). For example, it is common to think of a single bird as an individual, and not as a huge community of cells each doing their own thing.

Major transitions are important because the new higher-level organisms that they produce can lead to a great jump in complexity. For example, the evolution of multicellularity involved a transition from an entity with one part (the single-celled organism) working for the success of itself, to an entity with many parts (the multicellular organism), working for the success of the whole group. The cells can now have very different functions (a division of labour), as each is just a component of a multicellular machine, sacrificing itself for the good of the group, to get a sperm or egg cell into the next generation. As a result, diverse specialized forms such as eyes, kidneys, and brains were able to develop. The rise in complexity on Earth has been mediated by a handful of such jumps, when units with different goals (genes, single cells, individual insects) became intricately linked collectives with a single common goal (genomes, multicellular organisms, eusocial societies). Increases in complexity can also occur through mutations, gene duplications, or even whole genome duplications, but these are not major transitions. These other changes tend to be reversible and gradual, while major transitions are irreversible and cause large leaps in complexity.

The identification of major evolutionary transitions was an empirical observation about how complexity has increased on earth (Maynard Smith & Szathmáry 1995). The next step was to use evolutionary theory to provide insight about when (or under what conditions) we can expect major transitions to occur (Maynard Smith & Szathmáry 1995; Queller 1997; Gardner & Grafen 2009; Bourke 2011; West *et al.* 2015). Major transitions involve the original entities completely subjugating their own interests for the interests of the new collective. This represents an incredibly extreme form of cooperation. Think of the skin or liver cells in your body sacrificing for your sperm or eggs, or the worker ants in a eusocial colony sacrificing for the queen. Evolutionary theory tells us what conditions lead to such extraordinary cooperation.

What conditions drive major transitions?

Consider a multicellular organism, such as yourself. Why don't your hand and heart cells try to reproduce themselves, as opposed to helping your sperm or egg cells? The answer involves genetic similarity or 'relatedness' (Hamilton 1964). Your hand cells contain the same genes as your sperm cells because they are clonal copies. A hand cell could in principle get the same fraction of its genes into the next generation (all of them) by either copying itself, or by helping copy the sperm cells. A similar phenomenon occurs in eusocial insects, such as some ants, bees, wasps and termites. A worker termite can pass on half her genes to her offspring. But a random sibling in the colony (her brother or sister) also contains, on average, half her genes. Thus, a worker can get the same fraction of gene copies into the next generation by reproducing or by helping her mother, the queen, to reproduce (Hamilton 1964; Boomsma 2009). Helping their mother is likely to be more efficient than reproducing on their own, and so our termite can better get their genes into the next generation by helping rather than reproducing (Hamilton 1964; Queller & Strassmann 1998; Bourke 2011).

These are two examples of *alignment of interests*. The 'interests' are evolutionary interests in getting genes into future generations. The hand and the sperm cells both act as if they 'want' to get copies of their genes into the next generation, because as we discussed above, natural selection will have led to them being adapted in this way (Grafen 2003; Gardner 2009). The interests between them are aligned because they share the same genes. When individuals share genes, we say that they are genetically

related. Relatedness is a statistical measure of the extent to which individuals share genes (Grafen 1985).

In the case of eusocial ant colonies and human bodies, the interests are aligned through genetic relatedness. But there are other ways for evolutionary interests to be aligned. Consider, for example, a mutualism between two species. Some aphids carry bacteria in their gut (Moran 2007). The aphids provide the bacteria with sugars and other nutrients to survive and the bacteria provide the aphids with vital amino acids missing from their diet. The aphid and the bacteria do not share the same genes, but neither can reproduce without the other. To reproduce itself, the aphid has to help reproduce the bacteria and vice versa. Again, their evolutionary interests are *aligned*.

The very cells that make up our bodies – known as eukaryotic cells – evolved through a similar kind of alignment of interests (Margulis 1970; Thiergart *et al.* 2012; Archibald 2015). Early in the evolution of life, one bacterial species engulfed another. Over time, the two species took on different roles, with one specializing in replication and the other in energy production. The nucleus of our cells is the descendant of the former, and the mitochondria the latter. Neither can reproduce without the other. Their interests are aligned through reproductive dependence on each other.

All cooperation in nature requires alignment of interests (West *et al.* 2007). Consider, for example, flower pollination by bees. The bee benefits by receiving food from the flower, and the flower benefits by being pollinated. But major transitions are a particularly *extreme* form of cooperation. Compare the pollination scenario to the cells *within* the flower or the bee. Major transitions involve organisms cooperating so completely that they give up their status as individuals, becoming parts of a whole (Queller & Strassmann 2009). Unsurprisingly, then, major transitions require the extreme condition of *effectively* complete or perfect alignment of interests (Gardner & Grafen 2009; West *et al.* 2015).

It is also useful to consider the biology of organisms that do not have interests sufficiently aligned, and thus where conflict remains and major transitions have not occurred. For example, in single-celled organisms, we can compare non-clonal cooperative groups of things like slime moulds with clonal groups such as those that make up multicellular organisms such as humans and trees. These non-clonal groups have evolved only relatively limited division of labour, and never complex multicellular organisms (Fisher *et al.* 2013). Numerous experimental studies have shown that this is because in non-clonal groups non-cooperative 'cheats' can spread, limiting the extent of cooperation (Griffin *et al.* 2004; Diggle *et al.* 2007; Kuzdzal-Fick *et al.* 2011; Rumbaugh *et al.* 2012; Pollitt *et al.* 2014; Popat *et al.* 2015; Inglis *et al.* 2017).

Thus, there must be something in place to maintain the alignment of interests (Bourke 2011; West *et al.* 2015). Evolutionary theory can suggest what these somethings would have to be. In multicellular organisms, the something is the single-celled bottleneck (Buss 1987; Queller 2000). Multicellular organisms start each new generation as a single-celled zygote, such that all the cells in the resulting body are clonal (it could also be a spore giving rise to a haploid cell). Eusocial insect colonies evolved from colonies founded by a singly mated queen (Boomsma 2007, 2009, 2013; Hughes *et al.* 2008). If the queen had multiple mating partners, a worker would have half-sisters, and be less related to her siblings than her offspring, breaking down the alignment. The monogamous mating pair is the eusocial colony's equivalent of a zygote or a bottlenecking event (Boomsma 2013). With unrelated units, like mitochondria and the nucleus, the individual parts must be co-dependent for joint reproduction (Foster & Wenseleers 2006; West *et al.* 2015) – which can be thought of as a different form of bottleneck. The rarity of conditions like these – conditions under which alignment is so complete – explains the rarity of major transitions in individuality in the history of life.

Biology of organisms that have undergone major transitions

Do the conditions required for major transitions tell us anything about the biology of organisms that have undergone major transitions? Yes. Organisms are a nested hierarchy, where each nested level is the vestige of a former individual (Fig. 3). Eusocial ant colonies function as a single individual, but are made up of multicellular organisms. Those organisms themselves are made up of cells. In turn, those cells resulted from the fusion of two simple species early in evolution. Each of those organisms had a genome that evolved from the union of the individual, replicating molecules.

Further, at each level of the hierarchy, there must be something to *align the interests* of the parts. This usually happens through some form of population bottlenecking. When the parts are related, it is a relatedness bottleneck, such as the single-celled stage in multicellular organisms, or the singly mated female in the social insects (Boomsma 2009, 2013; West *et al.* 2015). When the parts are unrelated, it is usually another form of a bottleneck, such as enforced vertical transmission with joint reproduction (Foster & Wenseleers 2006; West *et al.* 2015). We use the term 'bottleneck' to refer to new generations being founded by a strict unit (the zygote, the mutualist pair, etc.), but another way to think of this is that the parts require each other for reproduction (e.g. the soma and the germ line, or the mitochondria and the nucleus). Other, further aligners may be required (e.g. in multicellular organisms, there may need to be a cap on somatic mutations), but these are more likely to be life-form specific.

To conclude so far, empirical observation tells us that complexity has increased on earth through major transitions. Evolutionary theory tells us that for major transitions to occur, the conflict must be eliminated. The theory also tells us what conditions lead to the elimination of conflict. The empirical data agree with the predictions of the theory, in that major transitions have only occurred in the extreme conditions that effectively remove conflict (Boomsma 2007; Hughes *et al.* 2008; Fisher *et al.* 2013; West *et al.* 2015; Fisher *et al.* 2017).

Complex aliens

Complexity and major transitions in space

We can now ask: what does the major evolutionary transition approach tell us about aliens? Will extraterrestrial life undergo major transitions? Not necessarily. Natural selection cannot predict a specific course of evolution. However, as we have said, we might be particularly interested in *complex* aliens. Complexity requires different parts or units working together towards a common goal or purpose. Under natural selection, units are selected to be selfish, striving to replicate themselves at the expense of others. Theory tells us that for units to unite under a common purpose, the evolutionary conflict between them must effectively eliminate (Gardner & Grafen 2009; West *et al.* 2015).

Once again, picture an alien (Fig. 2). If you are picturing something like unlinked replicating molecules or undifferentiated blobs

of slime, then your aliens might not have undergone major transitions. But if what you are picturing has different parts with specialized functions, then your alien is likely to have undergone major transitions (Fig. 2c). What matters is not that we call them 'major transitions', but rather that complexity requires multiple parts of an organism striving to the same purpose, and that theory predicts that this requires restrictive conditions (Gardner & Grafen 2009; West *et al.* 2015). Consequently, if we find complex organisms, we can make predictions about what they will be like.

Are there other ways to get complexity? To do so, natural selection would have to sculpt separate parts with unique functions out of a single replicator. Could, for example, the alien equivalent of a single copy of a gene, housed in one 'cell' generate the equivalent of limbs and organs? If so, it would disprove our prediction. However, both empirical (major transitions are how complexity has increased on Earth) and theoretical (functional parts requires the elimination of conflict) evidence support the argument that complex aliens will have undergone major transitions.

The biology of complex aliens

Given that complex aliens will have undergone major transitions, we can make a number of predictions about their biology (Fig. 4).

1. They will be entities that are made up of smaller entities – a nested hierarchy of individuality with as many levels as completed transitions. This could mean a collection of replicators, like the first genomes on the Earth, or some hideously complex nesting of groups on a planet where many more transitions have occurred than on our own. For example, you might imagine a 'society of societies', where many different social colonies collaborate, with each society specializing on different tasks, such that they are completely dependent on each other. Versions of the simpler entities are likely to be found free-living on the planet as well.
2. Whatever the number of transitions, there will be something that aligns interests, or eliminates conflict within the entities, at the level of each transition.

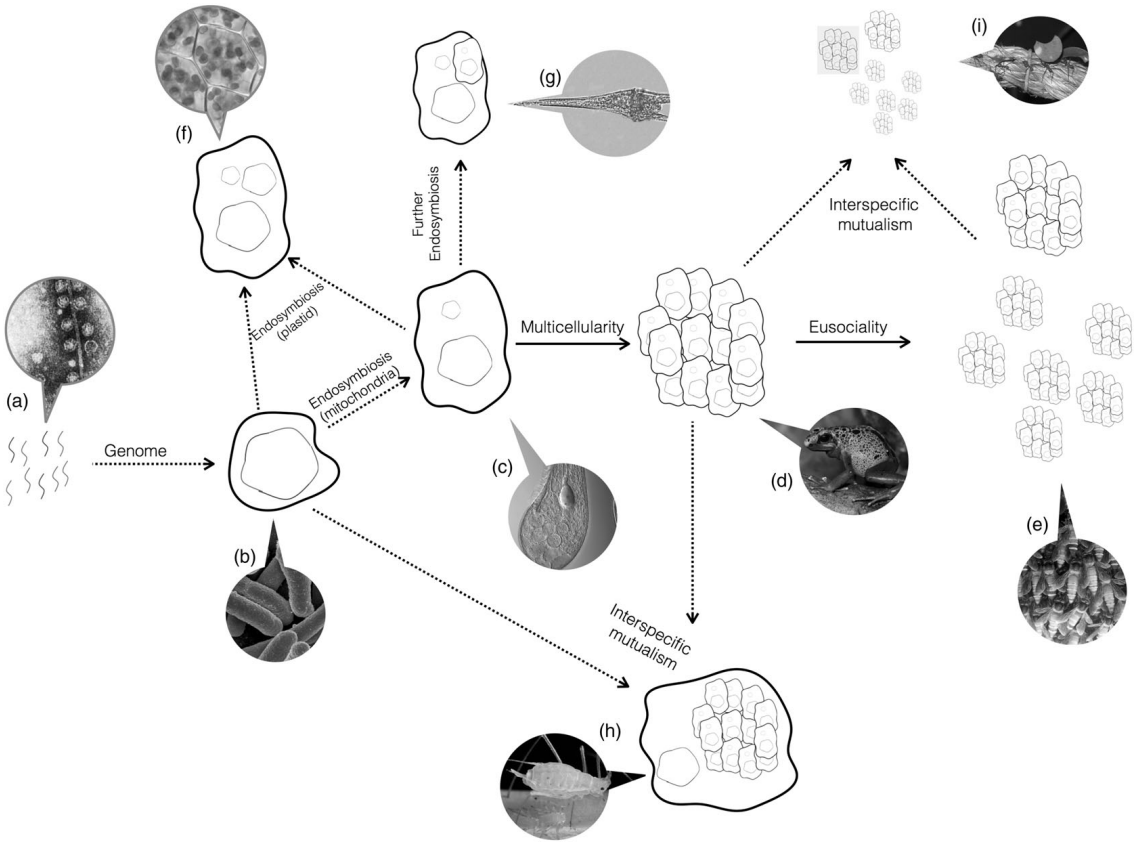

Fig. 3 - B/W online

Fig. 3. Major Transitions. Life started with naked replicating molecules, and has since undergone a series of major transitions. Arrows show the occurrence of major transitions in individuality. Dotted arrows represent transitions between dislike things and solid lines represent transitions between like things. Callouts show examples of the present-day organisms that have undergone that transition but no further ones. (a) As we have not yet identified the earliest replicators, Spiegelman's monster, a simple replicating RNA molecule, is shown as an example candidate. (b) A single-celled bacteria, such as *Escherichia coli*. (c) A single-celled eukaryote, like *Blepharisma japonicum*. (d) A multicellular organism, like frogs. (e) An obligate eusocial colony, such as honeybees. (f) Secondary endosymbiosis events, such as the origin of the chloroplast. (g) Further endosymbiosis events, such as those leading to Dinoflagellates. (h) Obligate interspecific mutualisms, such as between aphids and buchnera bacteria. (i) Obligate mutualisms between a multicellular organism and eusocial colony, such as between leaf-cutter ants and fungi. All images courtesy of Wikipedia.

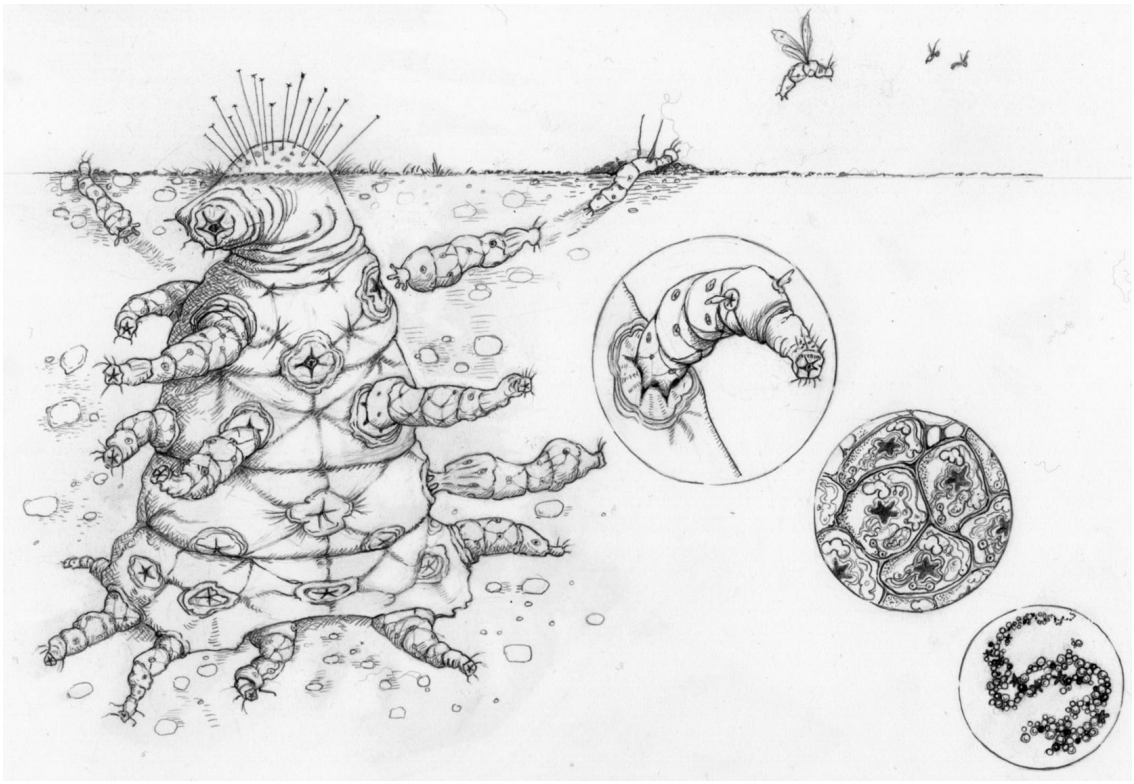

Fig. 4 - B/W online

Fig. 4. Major transitions in space: 'The Octomite'. A complex alien that comprises a hierarchy of entities, where each lower-level collection of entities has aligned evolutionary interests such that conflict is effectively eliminated. These entities engage in a division of labour, with various parts specializing on various tasks, such that the parts are mutually dependent.

3. Theory suggests that some sort of population bottlenecking will be key to aligning interests. Bottlenecking is not necessarily the only way to eliminate conflict, but it is probably the easiest evolutionary route to take. In particular, it does not require additional mechanisms of enforcement, such as kin discrimination, policing or randomization. The specific kinds of bottlenecking will depend on whether like or dislike units are united.
    a. When like entities come together, interests can be aligned through a bottleneck similar to our single-celled bottleneck in multicellular organisms or the single mating pair in eusocial colonies, which maximizes relatedness between entities.
    b. If the organisms are made up different types of entities, we can expect something similar to the bottleneck that forces mitochondria and nuclei to pass to the next generation together, with joint reproduction. By trapping individuals together over evolutionary time, their interests become aligned.
    c. Some aliens, like us, may contain both types of conflict reduction, for having both like and dislike types joined within them.

Conclusion

When using evolutionary theory to make predictions about extraterrestrial life, it is important to avoid circularity. Our chain of argument is: (1) Extraterrestrial life will have undergone natural selection. (2) Knowing that aliens undergo natural selection, we can make further predictions about their biology, based on the theory of natural selection. In particular, we can say something about complex aliens – that they will likely have undergone major transitions. (3) Theory tells us that restrictive conditions, which eliminate conflict, are required for major transitions. (4) Consequently, complex aliens will be composed of a nested hierarchy of entities, with the conditions required to eliminate conflict at each of those levels.

When making predictions about aliens, we must take advantage of our entire scientific toolkit. Mechanistic understanding is a good way to extrapolate from what we see on Earth. The theory is a good way to make predictions that are independent of the details of the Earth. Combining both approaches is the best way to make predictions about the many hundreds, thousands or millions of hypothetical aliens. Now we just need to find them.

Acknowledgements. We thank The Clarendon Fund, Hertford College, and the Natural Environment Research Council for funding; and Magdalen College for emergency housing.

Author disclosure statement. No competing financial interests exist.

References

Archibald JM (2015). Endosymbiosis and Eukaryotic Cell Evolution. *Current biology: CB* **25**(19), R911–921.

Benner SA (2003). Synthetic biology: Act natural. *Nature* **421**(6919), 118.
Boomsma JJ (2007). Kin selection versus sexual selection: why the ends do not meet. *Current biology : CB* **17**(16), R673–R683.
Boomsma JJ (2009). Lifetime monogamy and the evolution of eusociality. Philosophical transactions of the Royal Society of London. *Series B, Biological sciences* **364**(1533), 3191–3207.
Boomsma JJ (2013). *Nature's True Self. Science* (New York, N.Y.) **340**(6129), 145–146.
Bourke AFG (2011). *Principles of Social Evolution*. Oxford University Press.
Buss LW (1987) *The Evolution of Individuality*. Princeton University Press.
Cassan A, Kubas D, Beaulieu J-P, Dominik M, Horne K, Greenhill J, Wambsganss J, Menzies J, Williams A, Jørgensen UG, Udalski A, Bennett DP, Albrow MD, Batista V, Brillant S, Caldwell JAR, Cole A, Coutures C, Cook KH, Dieters S, Prester DD, Donatowicz J, Fouqué P, Hill K, Kains N, Kane S, Marquette J-B, Martin R, Pollard KR, Sahu KC, Vinter C, Warren D, Watson B, Zub M, Sumi T, Szymanski MK, Kubiak M, Poleski R, Soszynski I, Ulaczyk K, Pietrzynski G and Wyrzykowski L (2012). One or more bound planets per Milky Way star from microlensing observations. *Nature* **481**(7380), 167–169.
Cleland CE and Chyba CF (2002) Defining 'life'. *Origins of Life and Evolution of the Biosphere* **32**(4), 387–393.
Clutton-Brock TH, Harvey PH and Rudder B (1977). Sexual dimorphism, socionomic sex ratio and body weight in primates. *Nature* **269**(5631), 797–800.
Cohen J and Stewart I (2001). Where are the dolphins? *Nature* **409**(6823), 1119–1122.
Corning PA and Szathmáry E (2015). "Synergistic selection": a Darwinian frame for the evolution of complexity. *Journal of theoretical biology* **371**: 45–58.
Darwin C (1859). *On the origins of species by means of natural selection*. London: Murray 247.
Darwin C (1871). *The descent of man, and selection in relation to sex*. By Charles Darwin. New York, D. Appleton and company.
Davies NB and Houston AI (1981). Owners and Satellites: The Economics of Territory Defence in the Pied Wagtail, Motacilla alba. *The Journal of Animal Ecology* **50**(1), 157.
Davies NB, Krebs JR and West SA (2012) *An Introduction to Behavioural Ecology*. John Wiley & Sons.
Davies PCW, Benner SA, Cleland CE, Lineweaver CH, McKay CP and Wolfe-Simon F (2009). Signatures of a shadow biosphere. *Astrobiology* **9** (2), 241–249.
Des Marais DJ, Nuth JA, Allamandola LJ, Boss AP, Farmer JD, Hoehler TM, Jakosky BM, Meadows VS, Pohorille A, Runnegar B and Spormann AM (2008). The NASA Astrobiology Roadmap. *Astrobiology* **8** (4), 715–730.
Diggle SP, Griffin AS, Campbell GS and West SA (2007). Cooperation and conflict in quorum-sensing bacterial populations. *Nature* **450**(7168), 411–414.
Domagal-Goldman SD, Wright KE, Adamala K, Arina de la Rubia L, Bond J, Dartnell LR, Goldman AD, Lynch K, Naud M-E, Paulino-Lima IG, Singer K Walter-Antonio M, Abrevaya XC, Anderson R, Arney G, Atri D, Azúa-Bustos A, Bowman JS, Brazelton WJ, Brennecka GA, Carns R, Chopra A, Colangelo-Lillis J, Crockett CJ, DeMarines J, Frank EA, Frantz C, de la Fuente E, Galante D, Glass J, Gleeson D, Glein CR, Goldblatt C, Horak R, Horodyskyj L, Kaçar B, Kereszturi A, Knowles E, Mayeur P, McGlynn S, Miguel Y, Montgomery M, Neish C, Noack L, Rugheimer S, Stüeken EE, Tamez-Hidalgo P, Imari Walker S and Wong T (2016). The Astrobiology Primer v2.0. *Astrobiology* **16**(8), 561–653.
Fisher RA (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.
Fisher RM, Cornwallis CK and West SA (2013) Group formation, relatedness, and the evolution of multicellularity. *Current Biology*, Cell Press, **23** (12), 1120–1125.
Fisher RM, Henry LM, Cornwallis CK, Kiers ET and West SA (2017) The evolution of host-symbiont dependence. *Nature Communications*, Nature Publishing Group, **8**, 15973.
Flores Martinez CL (2014). SETI in the light of cosmic convergent evolution. *Acta Astronautica* **104**(1), 341–349.
Foster KR and Wenseleers T (2006). A general model for the evolution of mutualisms. *Journal of Evolutionary Biology* **19**(4), 1283–1293.
Gardner A (2009). Adaptation as organism design. *Biology Letters* **5**(6), 861–864.
Gardner A and Grafen A (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology* **22**(4), 659–671.
Grafen A (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology* **2**, 28–89.
Grafen A (2003). Fisher the evolutionary biologist. *Journal of the Royal Statistical Society: Series D (The Statistician)* **52**(3), 319–329.
Griffin AS, West SA and Buckling A (2004). Cooperation and competition in pathogenic bacteria. *Nature* **430**(7003), 1024.
Hamilton WD (1964). The genetical evolution of social behaviour I and II. *Journal of theoretical biology* **7**(1), 1–52.
Horneck G, Walter N, Westall F, Grenfell JL, Martin WF, Gomez F, Leuko S, Lee N, Onofri S, Tsiganis K, Saladino R, Pilat-Lohinger E, Palomba E, Harrison J, Rull F, Muller C, Strazzulla G, Brucato JR, Rettberg P and Capria MT (2016). AstRoMap European Astrobiology Roadmap. *Astrobiology* **16**(3), 201–243.
Hughes WOH, Oldroyd BP, Beekman M and Ratnieks FLW (2008). Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* (New York, N.Y.) **320**(5880), 1213–1216.
Inglis RF, Ryu E, Asikhia O, Strassmann JE and Queller DC (2017). Does high relatedness promote cheater free multicellularity in synthetic life-cycles?. *Journal of Evolutionary Biology* **30**(5), 985–993.
Kuzdzal-Fick JJ, Fox SA, Strassmann JE and Queller DC (2011). High relatedness is necessary and sufficient to maintain multicellularity in Dictyostelium. *Science* **334**(6062), 1548–1551.
Margulis L (1970). Recombination of non-chromosomal genes in Chlamydomonas: assortment of mitochondria and chloroplasts? *Journal of theoretical biology* **26**(2), 337–342.
Moran NA (2007). Symbiosis as an adaptive process and source of phenotypic complexity. *Proceedings of the National Academy of Sciences* **104** Suppl 1 (Supplement 1), 8627–8633.
Morris SC (2003). The navigation of biological hyperspace. *International Journal of Astrobiology* **2**(2), 149–152.
Petigura EA, Howard AW and Marcy GW (2013). Prevalence of Earth-size planets orbiting Sun-like stars. *Proceedings of the National Academy of Sciences of the United States of America* **110**(48), 19273–19278.
Pollitt EJ, West SA, Crusz SA, Burton-Chellew MN and Diggle SP (2014). Cooperation, quorum sensing, and evolution of virulence in Staphylococcus aureus. *Infection and immunity* **82**(3), 1045–1051.
Popat R, Pollitt EJ, Harrison F, Naghra H, Hong KW, Chan KG, Griffin AS, Williams P, Brown SP, West SA and Diggle SP (2015) Conflict of interest and signal interference lead to the breakdown of honest signaling. Evolution. *The Society for the Study of Evolution* **69**(9), 2371–2383.
Queller DC (1997). Cooperators Since Life Began The Major Transitions in Evolution. John Maynard Smith , Eors Szathmary. *The Quarterly Review of Biology* **72**(2), 184–188.
Queller DC (2000). Relatedness and the fraternal major transitions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355** (1403), 1647–1655.
Queller DC and Strassmann JE (1998) Kin selection and social insects. *Bioscience*, Oxford University Press, **48**(3), 165–175.
Queller DC and Strassmann JE (2009). Beyond society: the evolution of organismality. *Philosophical Transactions of the Royal Society B: Biological Sciences* **364**(1533), 3143–3155.
Rothschild LJ (2009). Defining the envelope for the search for life in the Universe. Proceedings of the International Astronomical Union 5(H15): 697–698.
Rothschild LJ (2010). A powerful toolkit for synthetic biology: Over 3.8 billion years of evolution. *BioEssays* **32**(4), 304–313.
Rumbaugh KP, Trivedi U, Watters C, Burton-Chellew MN, Diggle SP and West SA (2012). Kin selection, quorum sensing and virulence in pathogenic bacteria. *Proceedings of the Royal Society B: Biological Sciences* **279**(1742), 3584.

Schneider D (2016). $100 million seti initiative starts listening for E.T. *IEEE Spectrum* **53**(1), 41–42.
Shostak S (2015). Searching for Clever Life. *Astrobiology* **15**(11), 949–950.
Smith JM and Szathmáry E (1995). The major evolutionary transitions. *Nature* **374**(6519), 227–232.
von Salvini-Plawen L and Mayr E (1977). *On the Evolution of Photoreceptors and Eyes*. Boston, MA, Springer US: 207–263.
Thiergart T, Landan G, Schenk M, Dagan T and Martin WF (2012). An evolutionary network of genes present in the eukaryote common ancestor polls genomes on eukaryotic and mitochondrial origin. *Genome Biology and Evolution* **4**(4), 466–485.
West S (2009). *Sex Allocation*. Princeton, Princeton University Press.
West SA, Fisher RM, Gardner A, and Kiers ET (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences* **112**(33), 10112–10119.
West SA, Griffin AS and Gardner A (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of evolutionary biology* **20**(2), 415–432.

9 Discussion

Summary of Results

The preceding chapters each contained their own discussion. Accordingly, in this chapter I briefly summarise the main results, highlight some general points that emerge, and discuss future directions.

Chapter 2: The evolution of cooperation in simple molecular replicators

In Chapter 2, I considered the evolution of cooperation early in life's history. Before the origin of the genome, life likely consisted of simple molecular replicators that could do little more than act like enzymes to copy themselves or each other. Acting as an enzyme requires folding in such a way that potentially prevents one from being replicated as a template by others. As a result, there is a trade-off between enzymatic activity and replication rate. All else being equal, we expect parasitic replicators, acting only as templates, to invade the population. Previous work has suggested that limited diffusion on surfaces, which is a plausible scenario for early life, could maintain cooperative enzymatic activity by clustering cooperators together (Boerlijst and Hogeweg, 1991, 1995; Cronhjort and Blomberg, 1997; Szabó et al., 2002; Sardanyés and Solé, 2007; Bianconi et al., 2013; Shay et al., 2015;

McCaskill et al., 2001). However, if replicator cooperation is being driven by the same forces that favour cooperation in higher organisms, i.e. kin selection, this presents a puzzle. A standard result in kin selection theory is that limited diffusion is not a sufficient force to favour cooperation (Taylor, 1992).

I developed social evolution models of a system of simple molecular replicators. (i) I showed that we can understand cooperation in simple replicators as being driven by kin selection favouring positive enzymatic activity between related, or identical replicators. (ii) I showed that, as expected, limited diffusion is not sufficient to favour the evolution of cooperation. (iii) I showed that, instead, both overlapping generations and limited diffusion are required for cooperation to evolve. This explains previous results, revealing implicit assumptions that were key to the success of cooperation in earlier models. It also highlights the biological importance of an unusual feature of replicator life-history: their extreme degree of overlapping generations. Finally, it forms links to an entire body of well-developed theory: social evolution. Such links simplify our understanding of life, limiting the need to invoke separate explanations for different phenomena.

Chapter 3: Kin selection in the RNA world

In Chapter 3, I extended the work of Chapter 2 to consider problems of cooperation in early life more generally. The leading hypothesis for the the first evolving system of living things is the RNA world hypothesis. This hypothesis posits that, in the beginning, life consisted solely of RNA molecules acting as both templates and replicases. Various steps in the RNA world required cooperation, including the enzymatic activity discussed above, the co-existence of different kinds of ribozymes, and ultimately cooperation between different types of replicators to form a genome. However, there has previously been no overarching framework for understanding these problems, and they have been treated on a case by case basis.

I incorporated RNA cooperation into a kin selection framework. I developed models to highlight the potential importance of key features of RNA biology. First, I showed that one key factor in the evolution of RNA cooperation is the degree

to which cooperators can receive the benefits of cooperation. Simple replicators might be somewhat unusual in the tree of life in that acting as a co-operator may prevent one from receiving cooperation. This is different from the usual distinction between whole-group and others-only cooperation (Pepper, 2000). Whole-group cooperation occurs when some fraction of the cooperator's benefits return to the cooperator itself (e.g. a public good). Others-only cooperation occurs when the benefits of the cooperator's action do not return to the cooperator (e.g. a donation of food), but the cooperator may still receive benefits from its social partner. The case I considered in Chapter 3, which is relevant to the RNA world, is when, by acting as a cooperator, the focal individual cannot receive benefits from anyone, including its social partner. I showed that the degree to which this is true is a major factor in determining the degree to which cooperation can evolve.

Second, I generalised the results of Chapter 2 to show that the scale of competition determines the degree to which cooperation can evolve in replicators. This is particularly important in an RNA world, in which the simple mechanisms that generate relatedness, such as limited diffusion, also generate local competition. RNA replicators will need additional mechanisms, such as overlapping generations or elastic environments, to escape such competitive effects (Lehmann and Rousset, 2010). Third, I re-derived the results of Chapter 2 as a specific case of this model, placing those results in the broader framework I developed. Finally, I used this framework to unify previously disparate work on questions of cooperation in the RNA world, suggesting directions for future empirical studies. A key point is that simple, streamlined models like the ones presented in this chapter can highlight important biological features of the RNA world, guiding empirical work.

Chapter 4: A social coevolution hypothesis for the origin of the genome

In Chapter 4, I developed and tested a new hypothesis for the origin of the genome. Chapters 2 and 3 focused on cooperation between replicators of the same type. But ultimately the origin of the genome required cooperation between

replicators of different types. Previous work has invoked either highly specialised population structures or the presence of a cell to explain such cooperation, neither of which may be justified.

I argued that a simpler alternative explanation is that selection itself can drive the associations needed to jump-start the evolution of the genome. I hypothesised that if one type of replicator can act as a cooperative enzyme, and the other can act to physically associate, or stick to, other replicators, these two traits might coevolve, leading to a system of physically linked cooperative replicators. I tested this hypothesis using theoretical models. I showed that, given some passive byproduct benefits between replicator types, the tendency to physically associate and the tendency to act as an enzyme to replicate others can coevolve, leading to higher levels of cooperation and association than when either evolves on its own.

These results make the evolution of the genome easier to explain. We do not need to invoke a cell, a potentially complex piece of biology, to explain the genome. Indeed, by showing how a primitive genome could arise from simple biology, the results potentially free us to invoke the genome to explain the cell. Further, by minimising the need to invoke highly specialised population structures, these results expand the range of environments where the genome could have evolved. Finally, they shift the focus of genome-origin questions from being about external features of the environment, something we may never know, to biological features of replicators, something we can, increasingly, explore in synthetic biology laboratories.

Chapter 5: Inclusive fitness is an indispensable approximation for understanding organismal design

In Chapter 5, I addressed criticisms of inclusive fitness. Inclusive fitness has long been criticised for its assumptions, most notably additivity, and on the grounds that other measures of fitness, such as mean offspring number, do a better job at predicting gene frequency change in a wider range of scenarios. These criticisms have been around for decades, but have grown more common in the last decade. I outlined five advantages inclusive fitness offers as a biological maximand, including predicting

gene frequency change, offering a design principle, helping interpret behaviour, easing empirical work, and providing a single explanation for behaviour across a wide range of scenarios. I provided conceptual arguments for why alternatives, such as mean offspring number, are less useful in all ways except predicting gene frequency change. Further, I provided a verbal argument for why probabilistic mixing of phenotypes means that even this one advantage of other measures falls away in the face of biologically realistic assumptions. Finally, I offered some practicalities for behavioural ecologists, in particular for how to monitor assumptions so that one can be aware of the scenarios in which inclusive fitness might lead astray. In doing so, I hope to have clarified some issues that have remained obscure in the literature, and to have offered some assurance to biologists in the continuation of their use of inclusive fitness.

Chapter 6: Extending the range of additivity in using inclusive fitness

In Chapter 6, I addressed recent mathematical models which claimed to show the failure of inclusive fitness maximisation. Hamilton's original model showed, under the assumption of additivity of fitness effects, that inclusive fitness increases due to the action of natural selection. The idea has long been criticised for that assumption, and two recent formal analyses claim to show that when this assumption is relaxed, and non-additivity is allowed, inclusive fitness is not maximised (Lehmann et al., 2015; Okasha and Martens, 2016). I: (i) showed that in both cases, the authors failed to analyse the correct inclusive fitness, as Hamilton defined it; (ii) illustrated how to mathematically capture the correct inclusive fitness in these two models, suggesting a more general approach to capturing inclusive fitness in population genetic models; (iii) showed that, under a set of biologically plausible assumptions, inclusive fitness is indeed maximised. The biologically plausible assumption is that organisms' strategies include all probabilistic mixtures of strategies, an argument presented verbally in Chapter 5. I discussed the potentially wide ranging applicability of this assumption in biology.

Taken together, these results suggest a wider range of applicability of inclusive fitness, illustrate how mathematical biologists might go about incorporating inclusive fitness in highly technical arguments, and provide formal support for the widespread use of inclusive fitness thinking in empirical work. More generally, the dialogue between mathematicians and empirical biologists seems to have grown increasingly disparate in recent years. I hope that this work provides a bridge, such that social biologists can understand why inclusive fitness has appeared to fail in some models, and mathematical biologists can more appropriately incorporate the relatively nuanced idea of inclusive fitness into their models.

Chapter 7: Honest signalling and the double counting of inclusive fitness

In Chapter 7, I addressed a recent paper which claimed that kin selection has no effect on the honesty of signalling in birds (Bebbington and Kingma, 2017). Caro et al. (2016) argued that divorce should decrease the honesty of signalling in baby birds, because it decreases their relatedness to future siblings. In a reply, Bebbington and Kingma (2017) argued that Caro et al. (2016) miscalculated inclusive fitness, and that the 'correct' inclusive fitness predicts indifference to divorce. I showed that Bebbington et al's (2017) verbal argument was wrong, committing an error known as double counting. I developed formal models to demonstrate this point. These predict that, under standard assumptions, baby birds should indeed adjust their behaviour according to divorce, and the data support this prediction. I argued that the alternative explanations proposed by Bebbington and Kingma (2017) are either invalid or not competing explanations. This highlights the importance of using the correct inclusive fitness when trying to understand, explain, and predict social behaviours. It also illustrates the importance of using formal models to clarify what can otherwise be murky or vague verbal arguments. Finally, it illustrates the importance of inclusive fitness theory as a framework. Verbal arguments can lead to different predictions, and the assumptions of mathematical models can be hard to parse. The simplicity and interpretive value of inclusive fitness, when measured

correctly, can tell us what behaviours to expect, therefore offering a short-cut to identifying potential red-flags or miscalculations.

Chapter 8: Darwin's aliens

In Chapter 8, I considered the universality of natural selection, inclusive fitness, and major transitions in individuality. Astrobiology, which aims to answer questions of fundamental interest to humanity, is a rapidly growing field, a major goal of which is to make predictions about aliens. Previous work has largely relied on a mechanistic understanding of chemistry and physics, or on extrapolating from convergent evolution on Earth, which suffers from a sample size problem, because data points on earth are not independent.

I developed a series of arguments about how we might incorporate evolutionary theory into this prediction-making toolkit. (i) I provided an argument for why aliens will undergo natural selection, something which has often been either ignored or taken for granted. (ii) I argued that this allows us to use evolutionary theory to make predictions about aliens, which are potentially independent of Earth specific details, as they arise from logical features of the algorithm of natural selection. (iii) I argued that if we are particularly interested in complex aliens, we might expect them to be the product of major transitions in individuality. This is because such major transitions are how individual units collaborate to form more complex units. (iv) I argued that, because such transitions require extreme conditions due to the nature of the adaptive process, this fact allows us to make further predictions about aliens' biological structure and evolutionary history. For example, complex aliens will likely be a nested hierarchy of units, where those units were previously independent organisms, with mechanisms to minimise evolutionary conflict between them. Taken together, these arguments suggest that we may know more about aliens than previously thought, and suggest a larger role for evolutionary theory in making astrobiological predictions that escape the sample-size problem.

Emerging Themes and Future Directions

Early evolution

The results presented in this thesis suggest that cooperation near the start of life can be firmly incorporated into an inclusive fitness framework, and that doing so can be helpful. Compared to later major transitions, origin of the genome research has been somewhat scattered. Some of the work has used the inclusive fitness thinking that underpins the major transitions framework, but ignores many of the relevant biological features of early life, such as the need for unrelated types or the simplicity of replicator biology (e.g. Frank, 1994). Alternatively, work that has been attuned to the biology of early life has not utilised inclusive fitness thinking, occasionally rediscovering problems that are well known in social evolution, or missing solutions that have been identified in that field (e.g. Shay et al., 2015; Boerlijst and Hogeweg, 1995). Minimising the need for separate explanations for disparate phenomena is a major goal of evolutionary biology, and science more generally. This does not invalidate previous work – a plurality of approaches can only be beneficial. It simply suggests that making the links to an overarching framework can be useful, too.

Specifically, Chapters 2-4 highlight the value of bringing the genome into the major transitions fold. Doing so illuminated the key life history details of simple replicators that are likely relevant to the evolution of cooperation, something that can be obscured in complex simulations that are not motivated by a heuristic framework. Further, it revealed that a unifying theme for cooperation at the start of life is the presence of associations between replicators. This suggests that we may not need to invoke a cell or a highly specialised population structure to achieve these associations, and pointed to an alternative solution: the biology of the replicators themselves. Future empirical work, then, can focus on what phenotypes are possible in such simple replicators.

Chapter 4 presented a new hypothesis for the origin of a *primitive* genome. Ultimately, the evolution and maintenance of the genome required complete mutual dependence between the replicators, or genes. A future step would be to model

this problem, in order to understand what factors could drive the evolution of such interdependence. Recent work on the evolution of division of labour suggests that kin selection models can predict the circumstances under which mutual dependence evolves (Cooper and West, 2018). This work considered related individuals, and an application to multiple replicator species would be of interest.

Finally, Chapter 4 suggested that we may not need to invoke a cell to explain the genome, as a primitive genome can evolve without one. This is satisfying, as it was previously unclear how a cell might evolve without being able to be produced by a complex collection of genes. The exciting possibility here is that, once the genome evolves, it is potentially free to produce a cell. This leaves open the question of how the cell and the genome might coevolve.

Inclusive fitness theory

Inclusive fitness theory has long been plagued by criticism, despite its empirical successes (Foster, 2009; Westneat and Fox, 2010; Davies et al., 2012; Queller, 2016; Bourke, 2011b). As these criticisms become increasingly mathematical, it can be hard for biologists to know what to make of the competing arguments. Chapters 5 and 6 suggest that inclusive fitness holds more widely than some mathematical biologists have suggested. They also highlight the importance of capturing inclusive fitness correctly, according to Hamilton's verbal definition (Hamilton, 1964). Chapter 7 confirmed that this latter problem still affects some areas of empirical biology as well.

A key point of these sections is that inclusive fitness is a powerful organising framework that aids social biology. Calling it a framework might imply that it is merely a conceptual aid. After all, a framework like the major transitions framework aids empirical work, makes links across the tree of life, and simplifies the task of explaining disparate phenomena. However, there is as of yet a single general model to predict when a major transition in individuality occurs. Inclusive fitness, then, deserves a special status as a framework, because in addition to

those conceptual attributes, it also can be used as a predictive tool across an impressively wide range of scenarios.

Chapter 6 suggests that this range is wider than some mathematical biologists have claimed. However, I proved inclusive fitness maximisation under a general population structure with restricted pairwise interactions, and under general interactions but with a restricted population structure. A more general treatment, demonstrating inclusive fitness maximisation under general population structures *and* interactions, would be of interest for future work.

Astrobiology

Astrobiology has the aim of answering questions that are of fundamental scientific, philosophical, and human interest. Are we alone in the universe? What features of life are universal? Where should we look for life in space, and how can we detect it? With advancing technology, it's becoming increasingly possible to design and launch probes and telescopes that have the potential to discover alien life. I've argued that evolutionary theory is a powerful tool for this search, providing predictions that are potentially independent of details of Earth. One example, provided in this thesis, concerns major transitions in space.

Future work could develop game theoretic models of alien strategies, including relevant behaviours such as dispersal and signalling. Doing so requires formalising questions about what conditions allow us to model something as an evolutionary agent, identifying a fully general maximand that we can use to predict stable strategies, and providing a mutation-selection-independent theory for the origin of adaptation. A promising area for this final question is in information theory. The link between natural selection and information theory has already been made, revealing that we can conceptualise natural selection as causing organisms to accrue information about their environment (Frank, 2012b). Can we incorporate information theoretic dynamics of natural selection into an optimisation programme, to identify an information theoretic maximand? If we can, we may fully free

our agent-based models from life from Earth, which would be a tremendous coup for the search for life elsewhere.

Concluding remarks

Two themes have emerged from this thesis. First, inclusive fitness is a powerful organising framework and predictive tool. It was largely developed with higher organisms in mind. But as the reach of behavioural ecology extends to ever stranger organisms, such as bacteria, viruses, and even RNA replicators, inclusive fitness offers a tool to understand these new systems. Incorporating them into an existing framework can avoid re-discovering old problems, expediently point to solutions to new questions, and provide a heuristic for judging results and guiding empirical work.

In the process, formal models can clarify thinking and avoid unnecessary mistakes. However, if these models become too estranged from biology, it becomes easier to lose sight of the motivating principles. Doing so can lead to leaving out assumptions which might greatly simplify the analysis, or to making simple errors, such as miscalculating inclusive fitness. Biology without rigour becomes lost, but rigour without biology becomes meaningless.

Second, applying inclusive fitness to the fringes can help elucidate the theory itself. Simple replicators might have high mutant rates, or their possible phenotypes might be far apart in phenotype-space. Aliens may not have DNA, or they may lack familiar physical boundaries between organisms. How do these exotic features impact natural selection? To what degree can we expect these organisms to appear designed as if to maximise their inclusive fitness? In studying systems where standard assumptions are violated, we are forced to consider the importance of these assumptions. This can expand or contract the theory's remit – either way the remit is sharpened.

Biologists are fortunate among scientists to be in possession of remarkably simple theories that can explain a vast range of phenomena. Foremost among these is the expanded view of natural selection, that which takes into account social interactions: inclusive fitness theory. Extending this theory to its edges, in time, theory, and space, is a worthwhile pursuit.

Appendices

A

Modeling relatedness and demography in social evolution

doi:10.1002/evl3.69

Modeling relatedness and demography in social evolution

Guy A. Cooper,[1,2,*] Samuel R. Levin,[1,3,*] Geoff Wild,[4] and Stuart A. West[1]

[1]Department of Zoology, University of Oxford, Oxford OX1 3PS, United Kingdom

[2]E-mail: guy.cooper@zoo.ox.ac.uk

[3]E-mail: samuel.levin@zoo.ox.ac.uk

[4]Department of Applied Mathematics, University of Western Ontario, London, Ontario N6A 3K7, Canada

Received February 19, 2018
Accepted June 18, 2018

With any theoretical model, the modeler must decide what kinds of detail to include and which simplifying assumptions to make. It could be assumed that models that include more detail are better, or more correct. However, no model is a perfect description of reality and the relative advantage of different levels of detail depends on the model's empirical purpose. We consider the specific case of how relatedness is modeled in the field of social evolution. Different types of model either leave relatedness as an independent parameter (open models), or include detail for how demography and life cycle determine relatedness (closed models). We exploit the social evolution literature, especially work on the evolution of cooperation, to analyze how useful these different approaches have been in explaining the natural world. We find that each approach has been successful in different areas of research, and that more demographic detail is not always the most empirically useful strategy.

KEY WORDS: **Closed models, demography, evolutionary theory, life cycle, modeling, open models, population structure, relatedness.**

Theoretical models are often used to help explain how organisms behave in the natural world (Westneat and Fox 2010; Davies et al. 2012). In the field of social evolution, we use theoretical models to make predictions about and to ultimately understand behaviors that affect the fitness of individuals other than the actor (Hamilton 1964; Frank 1998; Bourke 2011). For example, we use models to predict when it is advantageous for individuals to cooperate; we use models to uncover the factors that contribute to the origin of selfish, altruistic, and even spiteful behaviors; and we use models to account for variation in the tendency to help both within and between species.

Perhaps the most influential model in social evolution was proposed by Hamilton (1964) and showed that genetic relatedness can be a key factor in explaining the adaptive value of social behaviors. Genetic relatedness is the probability that a social partner shares the same gene at a given locus relative to that of a random individual sampled from the population (Hamilton 1964, 1970; Grafen 1985). In large outbreeding populations, full siblings are related by ½, half-sibs by ¼, and so on (Grafen 1985). Individuals are favored to help relatives as this provides an indirect opportunity to further spread identical copies of their genes into the next generation. Over the last 50 years, relatedness has proven to be a fundamental concept for explaining social behavior across the tree of life, and theoretical models employing genetic relatedness have formed a cornerstone of social evolution (Frank 1998; Rousset 2004; West 2009; Bourke 2011).

The way in which relatedness is captured in theoretical models can be divided into two approaches, termed "open" and "closed" models (Box 1) (Taylor and Frank 1996; Frank 1998; Rousset 2004; Gardner and West 2006; Lion et al. 2011). In an open model, relatedness is left as an independent parameter that can be directly tuned by the theoretician without affecting the other features of the model. In a closed model, the modeler goes an extra step, to make specific assumptions about how population structure and life cycle determine relatedness. For example, the modeler might specify how model parameters, such as dispersal from the natal patch, the extent to which generations overlap, or

*Joint first authors.

the degree of monogamous mating impact relatedness from one generation to the next.

A potential problem with open models is that relatedness is not necessarily an independent variable (Taylor 1992a, 1992b). The factors that determine relatedness can influence other important factors. For example, patterns of dispersal and whether generations overlap can affect both relatedness and the relative marginal costs and benefits of social traits. Consequently, assuming that relatedness is an independent parameter in an open model could give misleading predictions. In contrast, closed models can take account of how different parameters are correlated, and so could be argued to be more correct or internally consistent. Closed modeling has become the most common approach in the field of social evolution, and has been suggested as the preferable method (Lehman and Rousset 2010; Lion et al. 2011). This raises the question of whether open models should be used.

Our aim is to critically analyse the utility of both open and closed approaches. Our starting point is two propositions, which we presume are widely agreed upon: (1) All models are wrong, in that they are not an exact representation of the natural world. (2) The usefulness of any model is determined by its ability to help explain the natural world. These two points are trivially true, but there has been little guidance in the literature for empirically minded theoreticians on when to develop one type of model over the other. We first examine the theoretical trade-offs of each approach and consider how they may be appropriate for different empirical questions. We then consider a few areas where open and closed models have been developed, including cooperation, sex allocation, and dispersal. We evaluate the success of each approach in explaining empirical patterns in these areas, to see if any lessons can be drawn for future research.

BOX 1: Open and closed: A toy model

We develop a simple model of public goods, first with an open and then a closed approach, to illustrate the two methods. We model the most general form of a public good, following Hamilton (1964), Taylor (1992a, 1992b), and Frank (2010). We take an inclusive fitness approach because the fitness derivations are simpler in this case, though an equivalent direct (neighbor modulated) fitness approach can be found in Taylor et al. (2007) and Levin and West (2017b).

Open Model: Some organism, such as a microbe, produces some costly public good, the benefits of which are shared between its social partners and itself. Examples in nature of public goods include the production and release of molecules by bacteria that scavenge for iron or digest protein (Griffin et al 2004; Diggle et al 2007). Because the production of the public good is costly to the individual, we might expect natural selection to favor individuals that do not incur the cost of production, but reap the benefits of good-producing social partners. Thus, we are interested in the conditions that would favor the evolution of the public good producing trait.

We assume an infinite population of individuals subdivided into social groups of size N (the infinite island model). Individuals can produce the public good at some private fecundity cost, c, which provides some fecundity benefit, b, to all individuals on the patch (including the focal individual). Hamilton (1964, 1970) showed that a trait will spread if its inclusive fitness effect, W_{IF}, is greater than *0* ($W_{IF} > 0$), where the inclusive fitness effect of an actor's trait is its effect on all individuals in the population, weighted by relatedness of the actor to those affected individuals (including the actor itself), or "recipients." In this case, the trait has a negative cost to the actor (with relatedness 1), and the relatedness to recipients is *r*, the average whole group relatedness in a social group (as opposed to others-only relatedness). Thus, the trait will spread if:

$$rb - c > 0,$$

which is a simple form of Hamilton's (1964) rule with b and c as simple additive fitness effects, as opposed to the general, regression form of Hamilton's rule (Gardner et al. 2011b). This is an open model, in which the mechanism by which *r* is generated is undefined. Positive relatedness in this model could come about through limited dispersal, kin recognition, partner choice, or any other process that generates genetic correlations within social groups. However, if *r* is correlated with the other model parameters (*b* and *c*), the predictions of this model might not be very useful for explaining variation in nature.

Closed Model: We might, for example, be interested in the case in which relatedness is generated through limited dispersal. We can capture this by incorporating a new parameter, *d*, which measures the proportion of offspring that disperse from their natal social group (with a fraction *(1-d)* remaining in the group). Following Taylor (1992a), we must now take into account not only the offspring produced as a direct result of public goods production, but also those offspring indirectly displaced as a result of the cooperative trait. An individual that expresses the public good trait incurs a fecundity cost, *c*, with relatedness 1, and provides a fecundity benefit, *b*, to recipients whose average relatedness is *r*. These extra $(b - c)$ offspring remain in the social group with probability $(1 - d)$, in which case the individuals they displace are also native with probability $(1 - d)$, and therefore have relatedness r. The overall inclusive fitness effect, then, is

$$W_{IF} = rb - c - r(1 - d)^2(b - c).$$

The above is still an open model, assuming independence between relatedness and model parameters. This illustrates that in principle, up until this point open and closed models can incorporate the same amount of demographic detail (though in practice, open models often do not). Taylor (1988, 1992a) showed how we can close the model by making additional assumptions. Specifically, he calculated relatedness in terms of the demographic parameters of the model (d & N). We can do this by writing the following population genetic recursion for the change in relatedness in a social group from one generation to the next:

$$r_{t+1} = 1/N + r_t(1-d)^2(N-1)/N.$$

Where the first term is the chance that two randomly sampled individuals on the patch are the same individual, and have relatedness one, and the second term is the chance they are different individuals both native to the patch, and therefore have the relatedness from the previous generation. Solving for the equilibrium value of relatedness, and plugging into the inclusive fitness effect above, we find the condition for the trait to spread is:

$$b/N > c.$$

This is Taylor's classic result—that the dispersal rate has no impact on whether the trait will spread.

Extensions: we can extend this closed model a number of ways to look at the impact of different life histories and explicit demographic parameters (Table 2). We do this by rewriting the fitness function and recalculating our estimate of relatedness accordingly. As one example, Taylor and Irwin (2000) allowed for overlapping generations by including a parameter s, the probability that a parent survives into the next generation. The inclusive fitness effect becomes:

$$W_{IF} = (1-s)[(rb-c) - r(1-d)^2(b-c).$$

Plugging in the equilibrium relatedness value, calculated in terms of s, d, and N, the condition for the public good trait to evolve becomes:

$$b/c > N - (N-1)[(2s(1-d))/((2-d)(1+s))].$$

The Scale of Competition

Open models can be used to provide an alternate way to look at the factors that arise in closed models (Frank 1998, Gardner and West 2006). For example, Frank (1998) developed a model for incorporating competition into an open model, by subsuming the scale of competition into benefit term of Hamilton's rule:

$$RB - C > 0$$

Where $R = r$, C = c, and B = b – a(b – c), and a is the proportion of competition that happens locally.

Queller (1994) developed a similar approach in which competition is subsumed into the relatedness parameter:

$$RB - C > 0$$

Where B = b, C = c, and $R = (r - ar)/(1 - ar)$, and therefore relatedness is not to an average member of the population but to an average competitor. Both the Queller (1994) and Frank (1998) approaches recover Taylor's (1992a) result as a specific case (see Gardner and West (2006) for further discussion).

The Trade-offs of Open and Closed Models

Open and closed modeling approaches differ in how they treat relatedness. Across nature, there is a wide diversity of life cycles and demographic structures that can generate relatedness between interacting individuals (Hamilton 1964; Frank 1998; Rousset 2004). Some well-characterized examples include:

1. Kin discrimination–if individuals can somehow distinguish relatives from nonrelatives and preferentially direct cooperation toward them, then this can generate positive relatedness between actor and recipient (Sharp et al. 2005; Mehdiabadi et al. 2006).
2. Dispersal patterns–limited dispersal, or dispersing as groups of relatives, can keep relatives together and hence generate positive relatedness between interacting individuals, in the absence of any kin discrimination (Hamilton 1964).
3. Mating patterns–monogamy or lower levels of polyandry can increase the relatedness between interacting siblings (Boomsma 2007; Hughes et al. 2008; Cornwallis et al. 2010, 2017; Lukas and Clutton-Brock 2012a).

OPEN MODELS

An open model is agnostic about which of the above factors (or others) are responsible for the generation of relatedness between individuals. Instead, relatedness is deliberately left as an independent factor that can be tuned directly by the modeler. The benefit of this approach is that it can generate predictions that should hold across many systems, regardless of which specific demographic processes are responsible for relatedness between interacting individuals. Thus, if the model predicts that investment in a public good will increase for higher relatedness, then this should hold just as well in systems that employ kin discrimination, limited dispersal or monogamous mating in the generation of relatedness.

The downside of an open approach is that relatedness is not necessarily independent of other factors. For example, relatedness can be an important driver of the evolution of dispersal, but relatedness also crucially depends upon dispersal (Taylor 1988; Frank 1998). Open models miss such feedbacks (West et al. 2002; Lehmann and Rousset 2010). Consequently, open models may gain widespread applicability, but at a cost of demographic precision.

CLOSED MODELS

Closed models

In contrast, a closed model specifies the precise way in which population dynamical processes generate genetic relatedness (Table 2). In doing so, concrete assumptions must be made about the exact life cycle and demography of the system and how these factors contribute to the relatedness of interacting individuals.

The benefit of a closed-model approach is that it allows a specific question to be answered about a characterized system, in which the processes that generate relatedness are known. Any feedback effects between parameters or traits of the model with the underlying genotypic assortment in the population are captured by the model. Furthermore, because the population-genetic assumptions about relatedness are clearer, closed models lend themselves to tweaking and altering assumptions or parameters in a way that allows theoreticians to build a family of related models, for which the intermodel relationships are apparent (Table 2).

However, the final step of closing a model involves determining precisely how a specific demography generates relatedness. Consequently, any conclusions drawn might only be applicable to that or a limited number of scenarios. This gives a precise solution, but it might be precisely irrelevant to what occurs in the real world. In fact, the way that relatedness arises in natural systems is frequently not well understood, arising from a convoluted combination of factors and processes. As such, the additional demographic assumptions that make closed models solvable are sometimes so idealized that they may add less realism to the model than might otherwise be expected (Taylor 1992a, 1992b; Gardner and West 2006; Lehman and Rousset 2010; Table 2). Consequently, closed models gain precise demographic detail, but at a cost of broader applicability.

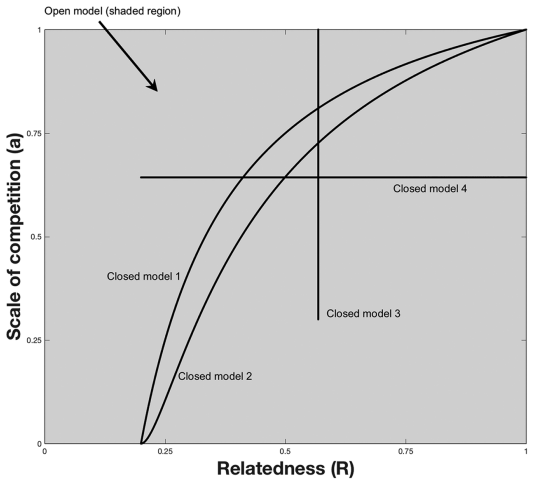

Figure 1. The relation between open and closed models. Frank (1998) developed an open model to show how local competition could reduce selection for cooperation between relatives. He used a parameter "*a*" to measure the scale at which density-dependent competition occurs, which can range from completely global (a = 0) to completely local (a = 1). In this figure, *a* is plotted against relatedness (*R*). Frank allowed these two variables to vary independently, and so his model encompasses the entire plane (shaded gray). In a closed model, we assume a specific demography and life history, and this causes *a* and *R* to be correlated in a specific way, leading to a particular curve in the plane (dark lines). For example, Closed model 1 is Taylor's 1992a model, closed model 2 is Taylor and Irwin's (2000) overlapping generations model, and closed models 3 and 4 are Gardner and West's 2006 budding dispersal model, for a fixed budding dispersal rate and range of migration rates, and a fixed migration rate and range of budding dispersal rates, respectively. Adapted from Gardner and West (2006).

OPEN VERSUS CLOSED

The differences between open and closed models can be illustrated graphically. Figure 1 graphs the relatedness (*R*) between interacting individuals versus the extent to which density dependent competition is at the scale of the local patch (*a;* Frank 1998). An open model can allow both these parameters to vary independently (the entire parameter space). A closed model determines how these parameters are related for a specified demography (one line on the figure). There are many different possible demographic scenarios and corresponding closed models (different lines on the figure). We provide some examples, which illustrate how different demographic assumptions can qualitatively change whether and how *R* and *a* are linked. This figure also illustrates how an open model can be used as a "meta-model" to examine how different closed models work and relate to each other (Frank 1998).

While there is a rough correlation between "open and closed" and "simple and complex," this is not always the case. In principle, closed models are nested within open models–up until the point of specifying relatedness, a closed model is open (Box 1). However, in practice, not all open models are one step away from being a closed model as the demography that determines relatedness and is required to close the model may not be specified at all (Wild 2011). Open models may instead include other ecological factors or otherwise unlinked demographic details and thus can be arbitrarily complex. Furthermore, in closed models, the interplay between different factors can sometimes lead to simpler predictions, as some parameters drop out of the analysis (Pen and Weissing 2000). Consequently, the difference between open and closed models may often be less of a distinction in complexity rather than a differing emphasis in the kinds of details that are included.

The above is a conceptual discussion of the relative trade-offs of open and closed modeling. However, the utility of different theoretical approaches is not a philosophical question, it is something that needs to be empirically tested. What matters is the interplay

between theory and data. Luckily, such an analysis is possible, via the extensive theoretical and empirical literature on the evolution of cooperation.

The Evolution of Cooperation: An Illustrative Example

A behavior or trait is defined as cooperation if it provides a benefit to another individual, and has evolved at least partially because of this benefit (West et al. 2007b). Cooperation poses an evolutionary problem because, all else being equal, it would reduce the relative fitness of the co-operator, and hence be selected against. There is a rich theoretical and empirical literature explaining the factors that can favor cooperation (Sachs et al. 2004; West et al. 2007a; Bourke 2011).

OPEN MODELS OF COOPERATION

A potential explanation for cooperation is that it is directed toward relatives, who also carry the gene for cooperation. By helping a relative reproduce, an individual is still passing copies of its genes to the next generation, just indirectly. This process, which is usually termed kin selection, was first modeled by Hamilton (1964) (Box 1). Hamilton showed that an altruistic cooperative trait will evolve if the fitness cost to the cooperator (C) is smaller than the fitness benefit (B) to the recipient, where the benefit to the recipient is weighted by the relatedness (R) of the cooperator to the recipient: $RB - C > 0$.

This result, known as Hamilton's rule, is an open model. Relatedness is a parameter (R) that is treated as independent of the other parameters of the model. There is no specification of how a positive R arises. As such, there are a number of population—and individual-level mechanisms that could generate a given R value.

Hamilton's rule has been employed to explain a wide range of traits across the tree of life (Table 1). It has been used to explain behavior, and variation in behavior, across diverse taxa, including bacteria, slime moulds, insects, birds, and mammals. The behaviors considered include many different forms of cooperation, policing, division of labor, dispersal, and harming behaviors such as killing or cannibalism. Furthermore, this includes cases where positive relatedness, or variation in relatedness, arises from a variety of factors, including limited dispersal, level of polyandry (promiscuity), kin discrimination and how groups are formed. In many cases, open models for more specific traits have also been developed (Table 1).

Closed models of cooperation

The open models discussed above black-boxed the mechanism that generated relatedness, and implicitly assumed that relatedness was independent of other model parameters. Over the last 30 years, many modelers interested in cooperation have instead employed closed models (Table 2).

Hamilton (1964) recognized that population viscosity via limited dispersal is a key mechanism for generating the positive relatedness values that can favor cooperation in Hamilton's rule. At the same time, however, limited dispersal can also increase competition between relatives, which reduces the relative benefit of helping relatives (Hamilton 1971, 1975). It is possible to put this local competition into an open model by adding an extra independent parameter or parameters (Grafen 1984; Frank 1998; Grafen and Archetti 2008). For example, RB-C-R_2D_2, where R_2 is the average relatedness between the actor and the individuals that suffer from increased competition and D_2 is the cost to these individuals (Grafen 1984). However, when parameters such as R and R_2 or B and D_2 are determined by the same factors, they will be correlated. Consequently, keeping them as independent parameters could give misleading predictions. For example, if limited dispersal increases both R and R_2, then we might not expect a higher relatedness (R) to lead to higher cooperation.

Taylor (1992a) developed a closed model of cooperation that considered the explicit effects of social group size and dispersal rates. He then estimated the value of relatedness as generated by the specific life-history details of the model. In a landmark result, he found that the dispersal rate had no influence on the evolution of cooperation. In Taylor's model, the effect of increased relatedness and competition exactly cancel. As such, Taylor's closed model predicted that a decrease in dispersal (and therefore an increase in relatedness) would not favor cooperation as predicted by the simple form of Hamilton's rule. As well as this specific result, for that exact life history, Taylor's model makes a general point about how we need to consider both cooperation and competition between relatives.

Taylor's model has since been expanded into a number of other closed models that tweak the life history in some manner (Table 2). In many of these cases, the specific life cycle allows limited dispersal to increase relatedness (R), without being exactly cancelled by a decreased benefit to relatives (B). Consequently, in these models, limited dispersal can favor cooperation. For example, Taylor and Irwin (2000) found that overlapping generations increase relatedness without inflating the costs of competition. This happens because there is a population-level mechanism (parent survival) for genetic associations to accrue in the absence of extra offspring remaining on the patch and competing (Box 1).

However, these closed models have had relatively little impact on our empirical understanding of specific biological cases. There is only one empirical example from the natural world where the data suggests that the influence of dispersal rates on relatedness and competition exactly cancel out–competition for mates between male fig wasps (West et al. 2001). The closed models stimulated experimental evolution studies in bacteria,

Table 1. **Examples of some of the phenomena where an open model approach (Hamilton's rule) has helped us understand biological phenomena.**

Taxa	Trait/Phenomena explained	Cause of variation in *R*	Empirical approach	More specific open models
Bacteria	Public goods (extracellular factors)	Dispersal pattern	Experimental evolution (Griffin et al. 2004)	Brown 1999; West and Buckling 2003; Dionisio and Gordo 2006; Frank 2010
Bacteria	Quorum sensing	Dispersal pattern	Experimental evolution (Diggle et al. 2007; Rumbaugh et al. 2012; Pollitt et al. 2014; Popat et al. 2015)	Brown and Jonstone 2001
Bacteria	Killing (bacteriocins)	Kin discrimination, dispersal pattern	Experimental (Inglis et al. 2009)	Gardner et al. 2004
Bacteria	Symbiotic benefit	Dispersal pattern (transmission)	Comparative (Fisher et al. 2017)	Frank 1996a
Birds and mammals	Cooperative breeding	Level of polyandry	Comparative (Cornwallis et al. 2010; 2017; Lukas and Clutton-Brock 2012a, 2012b)	Charnov 1981
Birds and mammals	Cooperation	Kin discrimination	Observational, experimental, comparative (Komdeur 1994; Russell and Hatchwell 2001; Griffin and West 2003; Komdeur et al. 2004; Sharp et al. 2005; Cornwallis et al 2009)	–
Fungus	Cooperation	Group formation, kin discrimination	Experimental evolution (Bastians et al. 2016)	–
Insects	Eusociality	Level of polyandry	Comparative (Hughes et al. 2008)	Charnov 1978, 1981; Gardner et al. 2011a; Alpedrinha et al. 2013, 2014; Rautiala et al. 2014; Liao et al. 2015,
Insects	Policing	Level of polyandry	Experimental, Comparative (Wenseleers and Ratnieks 2006a, 2006b; Ratnieks et al. 2006)	Ratnieks 1988; Wenseleers et al. 2004a, 2004b
Insects	Killing	Haplodiploidy, dispersal pattern, kin discrimination	Observational, experimental (Grbic et al. 1992; Giron et al. 2004a, 2004b)	–
Insects	Reproductive restraint	Level of polyandry	Observational, comparative (Wensellers and Ratnieks 2004)	Wenseleers et al. 2003, 2004a
Salamanders	Cannibalism	Kin discrimination	Experimental (Pfennig and Collins 1993; Pfennig et al. 1994, 1999)	–

(*Continued*)

Table 1. Continued.

Taxa	Trait/Phenomena explained	Cause of variation in *R*	Empirical approach	More specific open models
Slime moulds	Fruiting bodies	Dispersal pattern, kin discrimination	Observational, experimental evolution, genomic (Mehdiabadi et al. 2006; Gilbert et al. 2007; Kuzdzal-Fick et al. 2011; Ostrowski et al. 2015; Noh et al. 2018)	–
Social groups of cells (across taxa)	Division of labor, sterile cells	Dispersal pattern	Comparative (Fisher et al. 2013)	Cooper and West 2018

Our list is illustrative, not exhaustive, and we provide examples of the consequences of variation in only a single parameter (*R*). More specific open models are often constructed for specific traits. In many cases, some form of Hamilton's rule emerges as a prediction and is useful for interpreting these models (Taylor and Frank 1996; Frank 1998). For some other traits, such as sex allocation, the results are still interpreted with kin selection, but Hamilton's rule per se is less useful for interpretation. Studies focusing on the consequences of variation in other parameters (B, C), and whether Hamilton's rule is satisfied, are reviewed elsewhere (Bourke 2011, 2014).

examining how patterns of dispersal can influence both relatedness and competition (Griffin et al. 2004, Kümmerli et al. 2009). However, these studies can be seen as "wet simulations" that validate theory, but do not actually measure the consequences of competition in nature. Further, the role of demographic details has been discussed but rarely tested in a number of taxa, including RNA replicators, birds, and killer whales (Hatchwell 2009; Johnstone and Cant 2010; Croft et al. 2017; Levin and West 2017a).

OPEN VERSUS CLOSED

Why have open models been more useful for explaining specific empirical examples of cooperation? We suggest seven, nonmutually exclusive possibilities: (i) a closed model specifies a certain demography, narrowing the organisms to which it can be applied; (ii) closed models include an additional layer of demographic detail, which can make them more complex, and harder for empiricists to apply (or at least, they appear to); (iii) open models can offer intuitive heuristics, like Hamilton's rule, which can be applied broadly, generate simple predictions, and facilitate interpretation of results; (iv) open models make predictions in terms of *R*, which will often be a relatively easy parameter to measure; (v) open models disentangle causal effects in similar way to experiments that try to manipulate single factors while keeping everything else fixed; (vi) open models can focus on other biological details of potential interest, rather than demography (e.g., partner sanctions, or how cooperative benefits are shared; West et. all 2002; Cooper and West 2018); and (vii) there may not be enough two-way interactions between those developing the theory and those collecting the data.

The utility of the different approaches can also be illustrated by imagining a hypothetical scenario in which theoretical work on cooperation had started with Taylor's (1992a) closed model. In this case, we would have been left with the prediction that limited dispersal (higher relatedness) does not favor cooperation. Empirically this is clearly not the case, as limited dispersal appears to play a key role in favoring cooperation in a broad range of taxa (Table 1). But, at the same time, Taylor's model has been incredibly influential in its own right. The point is that Taylor's closed model was useful when discussed against an open model (Hamilton's rule). Hamilton's rule said relatedness matters, and it clearly does (Table 1). Taylor's model showed that, in certain cases, things could be more complicated as competition can reduce selection or even negate selection for cooperation between relatives. This helped us explain the data from fig wasps and stimulated experiments on bacteria (West et al. 2001; Griffin et al. 2004; Kümmerli et al. 2009), and led to a large body of theoretical work (Lehmann and Rousset 2010; Van Cleve and Lehman 2013; Van Cleve 2015; Peña et al. 2015). Furthermore, the combination of open and closed models in this area also spurred work on how local competition can favor spiteful harming behaviors (Gardner and West 2004; Gardner et al. 2004; 2007; Lehmann et al. 2006).

Beyond Cooperation

How useful have open and closed models been more generally? Another area of social evolution where there has been productive interplay between theory and data is the study of how organisms allocate resources to male and female offspring, termed sex

Table 2. Examples of the ways that Taylor's (1992a) model has been extended to incorporate additional biological details (nonexhaustive).

Theoretical models	Process modeled	When does limited dispersal favours cooperation?
Taylor 1992a	Patch elasticity	Always
Taylor and Irwin 2000, Irwin and Taylor 2001, Levin and West 2017b	Overlapping generations	When generations overlap
Gardner and West 2006, Lehmann et al. 2006, Lehmann et al. 2007, Traulsen and Nowak 2006	Budding dispersal	When individuals are more likely to disperse together than singly (budding).
Rogers 1990	Selective emigration	If altruists are more likely to emigrate
Gardner 2010, Johnstone and Cant 2008	Sex-specific dispersal	When the sex with higher variance in fitness is (slightly) more likely to disperse
Lehmann et al. 2008, Johnstone 2008	Caste-specific dispersal	When different castes (e.g. queen and worker) have different dispersal rates, reproductive values, and dispersal timings
Alizon and Taylor 2008	Empty sites	When there are empty sites on patches
El Mouden and Gardner 2008	Conditional helping	When co-operators adjust their behaviour conditional on whether they disperse
Taylor 1992b, Kelly 1992, Queller 1994, Gardner and West 2006	Various timings of cooperation and competition	Under some but not all demographic timing schemes
Yeh and Gardner 2012	Different ploidies	Under some but not all ploidies
Rodrigues and Gardner 2012, 2013a, b	Heterogeneity in patch quality, group size, and individual quality	When patches vary spatially and temporally in patch quality and group size, and (under some circumstances) when individuals vary in quality
Perrin and Lehmann 2001	Kin discrimination	When individuals can actively discriminate kin

We focus here on analytical models (rather than simulations), as these allow us to the see the explicit role of different parameters. We focus on island models, as opposed to spatially explicit models (e.g., lattice or stepping stone), as the added mathematical complexity of these models makes it harder to interpret parameter relationships, without necessarily revealing patterns that can't already be identified in simpler island models (Lehmann and Rousset 2010). A number of other models have used different approaches (e.g., lattice models, cellular automata, evolution on graphs) to identify a number of other factors that can alleviate the effects of local competition (e.g., van Baalen and Rand 1998; Mitteldorf and Wilson 2000; Ohtsuki et al. 2006; Lehmann et al. 2006; Grafen 2007; Taylor et al. 2007; Lion and Gandon 2009).

allocation (West 2009). Within this area, the two relevant success stories are: (1) local mate competition (LMC)–how population structuring, with competition for mates between related males, selects for female biased sex ratios (Hamilton 1967); (2) sex allocation driven by relatedness asymmetries in haplodiploid social insects (Trivers and Hare 1976; Boomsma and Grafen 1991). Closed and open models have driven research in these two areas respectively, demonstrating that, in different fields, one approach has sometimes been more useful than the other.

Hamilton (1967) showed that if n diploid females lay eggs on a patch, if mating then occurs on this patch, and if only the females disperse to compete globally, then the evolutionarily stable strategy is to invest a fraction $(n\text{-}1)/2n$ of resources into female offspring. The beauty of this closed model is that it is an excellent approximation of the life history of many species, and leads to a prediction in terms of one parameter that is often relatively easy to measure (n). A closed model works so well here, because clear morphological features, such as nondispersing wingless males, enforce life-history features that facilitate mathematical simplifications. Hamilton's LMC model has proved extremely useful for explaining variation in sex allocation, both within and between species (West 2009). Furthermore, theory has been extended in numerous directions to account for life history and demographic details relevant to certain species (West 2009). Alternative open formulations of Hamilton's LMC equation are possible, which focus on the relatedness between male and female offspring on a patch, but these can be less easy to apply (Frank 1998; Nee et al. 2002).

Boomsma and Grafen (1991) showed that, in haplodiploid social insects, workers are favored to adjust the colony sex allocation in response to the relatedness structure within their colony. They

produced an open model, and outlined how relatedness structure could be determined by a number of demographic factors, including queen mating rate, queen number, worker reproduction and queen replacement. Their model is able to explain considerable variation in sex allocation, between colonies (split sex ratios), in response to these factors (West 2009). A single open model could be applied across, and therefore unify, a number of different scenarios, where different features of the demography drive "split sex ratios." Together, these examples from sex allocation highlight that, for distinct empirical questions, different approaches have been more useful.

There are other areas where open or closed models have been more important for the development of theory. For example, closed models have dominated theoretical work on the evolution of dispersal, because the dispersal rate is both the trait under selection and the determinant of relatedness (Taylor 1988; Frank 1998; Gandon 1999; Gandon and Michalakis 1999; Gandon and Rousset 1999; Rousset 2004). Another example is the evolution of virulence, where early models tended to be open whereas later models are predominately closed (Frank 1996b; Gandon and Michalakis 2000; Wild et al. 2009; Alizon and Lion 2011; Lion 2013). However, neither of these fields has led to a similar interplay between theory and data, possibly because most of the theory was not developed to address specific empirical patterns (Crespi and Taylor 1990; Innocent et al. 2010).

Finally, there are also parameters other than relatedness that could be left open or closed. For example, in models where populations are structured into different classes—such as age, sex, or size—reproductive values are usually treated as closed. However, open models could be developed in these cases by employing a conservation of reproductive value criterion. Because total reproductive value of the population is constant, an increase in the reproductive value of one individual necessitates exact compensatory changes in the reproductive value of others, allowing the modeler to keep this as an open parameter (e.g., Wild and West 2007). Exactly how our analysis extends to these other questions remains unclear.

Guidelines

An obvious take home is that the different approaches have different utilities. But this is a bit vague and obvious. Can a summary of our above discussion provide more specific guidelines?

Open models have proved more useful when we want to consider cases where multiple demographic and life-history details can influence relatedness. For example, how limited dispersal, kin discrimination, and female mating rate influence the evolution of cooperation, or how queen mating rate, queen number, and queen replacement influence the evolution of split sex ratios (Hamilton 1964; Boomsma and Grafen 1991). In these cases, an open model can be applied broadly across diverse taxa, with very different life cycles. In addition, open models have been useful for providing conceptual unification, and intuitive heuristics for guiding empirical work.

Closed models have proved particularly useful when a single demographic factor is more universally important. For example, how the number of females laying eggs per patch influences sex allocation (Hamilton 1967). In such cases, a closed model can be applied broadly across different taxa, which share this key aspect of their life cycle. In addition, closed models have been useful conceptually for disentangling the roles of different demographic parameters.

More generally, with all these considerations, the emphasis should always be on the interplay between theory and data, and how the theory will be used to help us explain the natural world. When developing theory, there are a number of empirically motivated questions to be asked. What aspect of the empirical data can't be explained by existing theory and needs a new model? What are the parameters that empirical work suggests need more attention? Do we want to make broad predictions across species with different life cycles, or for a single species with a specific life cycle? The advantage of more empirically minded development of theory is clearly illustrated by the success of closed models developed to examine sex allocation (local mate competition), compared to those for cooperation and dispersal. In particular, the extensions of basic local mate competition theory have proven very useful precisely because their development was driven by cases where the data and/or life-history assumptions did not fit existing theory (West 2009).

Conclusions

To conclude, open and closed models are complementary and not competing approaches. Ultimately, we must ask what the modeler is prepared to give up, and what they want to gain, which will depend on the modeler's empirical aim. Sylvain Gandon pointed out to us that an analogy here can be made with the analysis of statistical data. If the addition of an extra variable does not significantly improve the explanation of the data, then the more detailed model, with that extra variable, can be a less good model, as judged by statistical measures such as AIC. An important goal should be to develop a model with the minimal level of detail required to answer a specific biological question (May 2004). Evaluating whether to use an open or closed model is then simply a matter of determining where that minimal level of detail falls with respect to demography and population structure.

Finally, this debate touches on a recurring theme in behavioral and evolutionary ecology, where there are numerous examples of different potential approaches. Some examples include

population genetics versus game theory, general versus specific models in game theory, or experimental studies on a specific species versus across species comparative studies (Harvey and Purvis 1991; Parker and Maynard Smith 1990; Davies et al. 2012). All of these cases have generated arguments that one approach is "better" or "more correct" than the other whereas, in reality, the different methodologies have different strengths and weaknesses and are each appropriate in different scenarios.

AUTHOR CONTRIBUTIONS

All authors contributed to the manuscript equally.

ACKNOWLEDGMENTS

We thank Sylvain Gandon, Miguel dos Santos, Andy Gardner, Michael Cant, and one anonymous reviewer for their helpful comments and discussion.

LITERATURE CITED

Alizon, S., and P. Taylor. 2008. Empty sites can promote altruistic behavior. Evolution 62:1335–1344.

Alizon, S., and S. Lion. 2011. Within-host parasite cooperation and the evolution of virulence. Proc. R Soc. Lond. B Biol. Sci. 278:3738–3747.

Alpedrinha, J., A. Gardner, and S. A. West. 2014. Haplodiploidy and the evolution of eusociality: worker revolution. Am. Nat. 184:303–317.

Alpedrinha, J., S. A. West, and A. Gardner. 2013. Haplodiploidy and the evolution of eusociality: worker reproduction. Am. Nat. 182:421–438.

Bastiaans, E., A. J. Debets, and D. K. Aanen. 2016. Experimental evolution reveals that high relatedness protects multicellular cooperation from cheaters. Nat. Comm. 7:11435.

Boomsma, J. J., and A. Grafen. 1991. Colony-level sex ratio selection in the eusocial Hymenoptera. Journal of Evolutionary Biology 4:383–407.

Boomsma, J. J. 2007. Kin selection versus sexual selection: why the ends do not meet. Curr. Biol. 17:R673–R683.

Bourke, A. F. 2011. Principles of social evolution. Oxford Univ. Press, Oxford.

Bourke, A. F. 2014. Hamilton's rule and the causes of social evolution. Phil. Trans. R Soc. B 369:20130362.

Brown, S. P. 1999. Cooperation and conflict in host–manipulating parasites. Proc. R. Soc. Lond. B Biol. Sci. 266:1899–1904.

Brown, S. P., and R. A. Johnstone. 2001. Cooperation in the dark: signalling and collective action in quorum-sensing bacteria. Proc. R Soc. Lond. B Biol. Sci. 268:961–965.

Charnov, E. L. 1978. Evolution of eusocial behavior: offspring choice or parental parasitism? J. Theoret. Biol. 75:451–465.

Charnov, E. L. 1981. Kin selection and helpers at the nest: effects of paternity and biparental care. Anim. Behav. 29:631–632.

Cooper, G. A., and S. A. West. 2018. Division of labour and the evolution of extreme specialisation. Nat. Ecol. Evol. 2:1161–1167.

Cornwallis, C. K., C. A. Botero, D. R. Rubenstein, P. A. Downing, S. A. West, and A. S. Griffin. 2017. Cooperation facilitates the colonization of harsh environments. Nat. Ecol. Evol. 1:0057.

Cornwallis, C. K., S. A. West, K. E. Davis, and A. S. Griffin. 2010. Promiscuity and the evolutionary transition to complex societies. Nature 466:969.

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

Crespi, B. J., and P. D. Taylor. 1990. Dispersal rates under variable patch density. Am. Nat. 135:48–62.

Croft, D. P., R. A. Johnstone, S. Ellis, S. Nattrass, D. W. Franks, L. J. Brent, et al. 2017. Reproductive conflict and the evolution of menopause in killer whales. Curr. Biol. 27:298–304.

Davies, N. B., J. R. Krebs, and S. A. West. 2012. An introduction to behavioural ecology. John Wiley & Sons, Hoboken, New Jersey.

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

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

El Mouden, C., and A. Gardner. 2008. Nice natives and mean migrants: the evolution of dispersal-dependent social behaviour in viscous populations. J. Evol. Biol. 21:1480–1491.

Fisher, R. M., C. K. Cornwallis, and S. A. West. 2013. Group formation, relatedness, and the evolution of multicellularity. Curr. Biol. 23:1120–1125.

Fisher, R. M., L. M. Henry, C. K. Cornwallis, E. T. Kiers, and S. A. West. 2017. The evolution of host-symbiont dependence. Nature Communications 8:15973.

Frank, S. A. 1996a. Host–symbiont conflict over the mixing of symbiotic lineages. Proc. R. Soc. Lond. B 263:339–344.

Frank, S. A. 1996b. Models of parasite virulence. Quart. Rev. Biol. 71:37–78.

Frank, S. A. 1998. Foundations of social evolution. Princeton Univ. Press, Princeton.

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

Gandon, S. 1999. Kin competition, the cost of inbreeding and the evolution of dispersal. J. Theoret. Biol. 200:345–364.

Gandon, S., and Y. Michalakis. 1999. Evolutionarily stable dispersal rate in a metapopulation with extinctions and kin competition. J. Theoret. Biol. 199:275–290.

Gandon, S., and Y. Michalakis. 2000. Evolution of parasite virulence against qualitative or quantitative host resistance. Proc. R Soc. Lond. B Biol. Sci. 267:985–990.

Gandon, S., and F. Rousset 1999. Evolution of stepping-stone dispersal rates. Proc. R Soc. Lond. B Biol. Sci. 266:2507–2513.

Gardner, A., and S. A. West. 2006. Demography, altruism, and the benefits of budding. J. Evol. Biol. 19:1707–1716.

Gardner, A. 2010. Sex-biased dispersal of adults mediates the evolution of altruism among juveniles. Journal of Theoretical Biology 262:339–345.

Gardner, A., J. Alpedrinha, and S. A. West. 2011a. Haplodiploidy and the evolution of eusociality: split sex ratios. The American Naturalist 179:240–256.

Gardner, A., S. A. West, and G. Wild. 2011b. The genetical theory of kin selection. Journal of Evolutionary Biology 24:1020–1043.

Gardner, A., I. C. Hardy, P. D. Taylor, and S. A. West. 2007. Spiteful soldiers and sex ratio conflict in polyembryonic parasitoid wasps. Am. Nat. 169:519–533.

Gardner, A., S. A. West, and A. Buckling. 2004. Bacteriocins, spite and virulence. Proc. R Soc. Lond. B Biol. Sci. 271:1529–1535.

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

Giron, D., D. W. Dunn, I. C. Hardy, and M. R. Strand. 2004a. Aggression by polyembryonic wasp soldiers correlates with kinship but not resource competition. Nature 430:676.

Giron, D., S. Pincebourde, and J. Casas. 2004b. Lifetime gains of host-feeding in a synovigenic parasitic wasp. Physiological Entomology 29:436–442.

Grafen, A. 1984. Natural selection, kin selection and group selection. Behavioural Ecology: An Evolutionary Approach 2:62–84.

Grafen, A., and M. Archetti. 2008. Natural selection of altruism in inelastic viscous homogeneous populations. Journal of Theoretical Biology 252:694–710.

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

Grafen, A. 2007. An inclusive fitness analysis of altruism on a cyclical network. J. Evol. Biol. 20:2278–2283.

Grbić, M., P. J. Ode, and M. R. Strand. 1992. Sibling rivalry and brood sex ratios in polyembryonic wasps. Nature 360:254.

Griffin, A. S., and S. A. West. 2003. Kin discrimination and the benefit of helping in cooperatively breeding vertebrates. Science 302:634–636.

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

Hamilton, W. D. 1964. The genetical theory of social evolution, I and II. J. Theoret. Biol. 7:1–52.

Hamilton, W. D. 1967. Extraordinary sex ratios. Science 156:477–488.

Hamilton, W. D. 1970. Selfish and spiteful behaviour in an evolutionary model. Nature 228:1218.

Hamilton, W. D. 1971. Geometry for the selfish herd. J. Theoret. Biol. 31:295–311.

Hamilton, W. D. 1975. Innate social aptitudes of man: an approach from evolutionary genetics. Biosocial Anthropol. 53:133–55.

Harvey, P. H., and A. Purvis. 1991. Comparative methods for explaining adaptations. Nature 351:619.

Hatchwell, B. J. 2009. The evolution of cooperative breeding in birds: kinship, dispersal and life history. Philos. Trans. R Soc. B Biol. Sci. 364:3217–3227.

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

Inglis, R. F., A. Gardner, P. Cornelis, and A. Buckling. 2009. Spite and virulence in the bacterium *Pseudomonas aeruginosa*. Proc. Natl. Acad. Sci. 106:5703–5707.

Innocent, T. M., J. Abe, S. A. West, and S. E. Reece. 2010. Competition between relatives and the evolution of dispersal in a parasitoid wasp. J. Evol. Biol. 23:1374–1385.

Irwin, A. J., and P. D. Taylor. 2001. Evolution of altruism in stepping-stone populations with overlapping generations. Theoret. Popul. Biol. 60:315–325.

Johnstone, R. A., and M. A. Cant. 2008. Sex differences in dispersal and the evolution of helping and harming. The American Naturalist 172:318–330.

Johnstone, R. A., and M. A. Cant. 2010. The evolution of menopause in cetaceans and humans: the role of demography. Proc. R Soc. Lond. B Biol. Sci.

Johnstone, R. A. 2008. Kin selection, local competition, and reproductive skew. Evolution 62:2592–2599.

Kell, D. B., A. S. Kaprelyants, and A. Grafen. 1995. Pheromones, social behaviour and the functions of secondary metabolism in bacteria. Trends Ecol. Evol. 10:126–129.

Kelly, J. K. 1992. Kin selection in density regulated populations. J. Theoret. Biol. 157:447–461.

Komdeur, J. 1994. The effect of kinship on helping in the cooperative breeding Seychelles warbler (*Acrocephalus sechellensis*). Proc. R Soc. Lond. B 256:47–52.

Komdeur, J., D. S. Richardson, and T. Burke. 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.

Kümmerli, R., A. Gardner, S. A. West, and A. S. Griffin. 2009. Limited dispersal, budding dispersal, and cooperation: an experimental study. Evolution 63:939–949.

Kuzdzal-Fick, J. J., S. A. Fox, J. E. Strassmann, and D. C. Queller. 2011. High relatedness is necessary and sufficient to maintain multicellularity in *Dictyostelium*. Science 334:1548–1551.

Lehmann, L., K. Bargum, and M. Reuter. 2006. An evolutionary analysis of the relationship between spite and altruism. J. Evol. Biol. 19:1507–1516.

Lehmann, L., and F. Rousset. 2010. How life history and demography promote or inhibit the evolution of helping behaviours. Philosophical Transactions of the Royal Society of London B: Biological Sciences 365:2599–2617.

Lehmann, L., N. Perrin, and F. Rousset. 2006. Population demography and the evolution of helping behaviors. Evolution 60:1137–1151.

Lehmann, L., L. Keller, S. West, and D. Roze. 2007. Group selection and kin selection: two concepts but one process. Proc. Natl. Acad. Sci. 104:6736–6739.

Lehmann, L., V. Ravigné, and L. Keller. 2008. Population viscosity can promote the evolution of altruistic sterile helpers and eusociality. Proc. R Soc. Lond. B Biol. Sci. 275:1887–1895.

Levin, S. R., and S. A. West. 2017a. Kin selection in the RNA world. Life 7:53.

Levin, S. R., and S. A. West. 2017b. The evolution of cooperation in simple molecular replicators. Proc. R Soc. B 284:20171967.

Liao, X., S. Rong, and D. C. Queller. 2015. Relatedness, conflict, and the evolution of eusociality. PLoS Biol. 13:e1002098.

Lion, S. 2013. Multiple infections, kin selection and the evolutionary epidemiology of parasite traits. J. Evol. Biol. 26:2107–2122.

Lion, S., and S. Gandon. 2009. Habitat saturation and the spatial evolutionary ecology of altruism. J. Evol. Biol. 22:1487–1502.

Lion, S., V. A. Jansen and T. Day. 2011. Evolution in structured populations: beyond the kin versus group debate. Trends in Ecology & Evolution 26:193–201.

Lukas, D., and T. Clutton-Brock. 2012a. Cooperative breeding and monogamy in mammalian societies. Proc. R Soc. Lond. B Biol. Sci. 279:2151–2156.

Lukas, D., and T. Clutton-Brock. 2012b. Life histories and the evolution of cooperative breeding in mammals. Proc. R Soc. Lond. B Biol. Sci. 279:4065–4070.

May, R. M. 2004. Uses and abuses of mathematics in biology. Science 303:790–793.

Mehdiabadi, N. J., C. N. Jack, T. T. Farnham, T. G. Platt, S. E. Kalla, G. Shaulsky, et al. 2006. Social evolution: kin preference in a social microbe. Nature 442:881.

Mitteldorf, J., and D. S. Wilson. 2000. Population viscosity and the evolution of altruism. J. Theoret. Biol. 204:481–496.

Nee, S., S. A. West, and A. F. Read. 2002. Inbreeding and parasite sex ratios. Proc. R Soc. Lond. B Biol. Sci. 269:755–760.

Noh, S., K. S. Geist, X. Tian, J. E. Strassmann, and D. C. Queller. 2018. Genetic signatures of microbial altruism and cheating in social amoebas in the wild. Proceedings of the National Academy of Sciences 201720324.

Ohtsuki, H., C. Hauert, E. Lieberman, and M. A. Nowak. 2006. A simple rule for the evolution of cooperation on graphs and social networks. Nature 441:502.

Ostrowski, E. A., Y. Shen, X. Tian, R. Sucgang, H. Jiang, J. Qu, et al. 2015. Genomic signatures of cooperation and conflict in the social amoeba. Curr. Biol. 25:1661–1665.

Parker, G. A., and J. Maynard Smith. 1990. Optimality theory in evolutionary biology. Nature 348:27.

Pen, I., and F. J. Weissing. 2000. Towards a unified theory of cooperative breeding: the role of ecology and life history re-examined. Proc. R Soc. Lond. B Biol. Sci. 267:2411–2418.

Peña, J., G. Nöldeke, and L. Lehmann. 2015. Evolutionary dynamics of collective action in spatially structured populations. J. Theoret. Biol. 382:122–136.

Perrin, N., and L. Lehmann. 2001. Is sociality driven by the costs of dispersal or the benefits of philopatry? A role for kin-discrimination mechanisms. Am. Nat. 158:471–483.

Pfennig, D. W., and J. P. Collins. 1993. Kinship affects morphogenesis in cannibalistic salamanders. Nature 362:836.

Pfennig, D. W., J. P. Collins, and R. E. Ziemba. 1999. A test of alternative hypotheses for kin recognition in cannibalistic tiger salamanders. Behav. Ecol. 10:436–443.

Pfennig, D. W., P. W. Sherman, and J. P. Collins. 1994. Kin recognition and cannibalism in polyphenic salamanders. Behav. Ecol. 5:225–232.

Pollitt, E. J., S. A. West, S. A. Crusz, M. N. Burton-Chellew, and S. P. Diggle. 2014. Cooperation, quorum sensing, and evolution of virulence in *Staphylococcus aureus*. Infection Immunity 82:1045–1051.

Popat, R., E. J. Pollitt, F. Harrison, H. Naghra, K. W. Hong, K. G. Chan, A. S. Griffin, P. Williams, S. P. Brown, S. A. West and S. P. Diggle. 2015. Conflict of interest and signal interference lead to the breakdown of honest signaling. Evolution 69:2371–2383.

Queller, D. C. 1994. Genetic relatedness in viscous populations. Evol. Ecol. 870–873.

Ratnieks, F. L. 1988. Reproductive harmony via mutual policing by workers in eusocial Hymenoptera. Am. Nat. 132:217–236.

Ratnieks, F. L., K. R. Foster, and T. Wenseleers. 2006. Conflict resolution in insect societies. Annu. Rev. Entomol. 51:581–608.

Rautiala, P., H. Helanterä, and M. Puurtinen. 2014. Unmatedness promotes the evolution of helping more in diplodiploids than in haplodiploids. Am. Nat. 184:318–325.

Rodrigues, A. M., and A. Gardner. 2012. Evolution of helping and harming in heterogeneous populations. Evolution 66:2065–2079.

Rodrigues, A. M., and A. Gardner. 2013a. Evolution of helping and harming in heterogeneous groups. Evolution 67:2284–2298.

Rodrigues, A. M., and A. Gardner. 2013b. Evolution of helping and harming in viscous populations when group size varies. Am. Nat. 181:609–622.

Rogers, A. R. 1990. Group selection by selective emigration: the effects of migration and kin structure. Am. Nat. 135:398–413.

Rousset, F. 2004. Genetic structure and selection in subdivided populations (MPB-40). Princeton Univ. Press, Princeton.

Rumbaugh, K. P., U. Trivedi, C. Watters, M. N. Burton-Chellew, S. P. Diggle, and S. A. West. 2012. Kin selection, quorum sensing and virulence in pathogenic bacteria. Proc. R. Soc. B 279:3584–3588.

Russell, A. F., and B. J. Hatchwell. 2001. Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. Proc. R Soc. Lond. B Biol. Sci. 268:2169–2174.

Sachs, J. L., U. G. Mueller, T. P. Wilcox, and J. J. Bull. 2004. The evolution of cooperation. Quart. Rev. Biol. 79:135–160.

Sharp, S. P., A. McGowan, M. J. Wood, and B. J. Hatchwell. 2005. Learned kin recognition cues in a social bird. Nature 434:1127.

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

Taylor, P. D., and A. J. Irwin. 2000. Overlapping generations can promote altruistic behavior. Evolution 54:1135–1141.

Taylor, P. D. 1988. An inclusive fitness model for dispersal of offspring. J. Theoret. Biol. 130:363–378.

Taylor, P. D. 1992a. Altruism in viscous populations–an inclusive fitness model. Evolutionary Ecology 6:352–356.

Taylor, P. D. 1992b. Inclusive fitness in a homogeneous environment. Proc. R Soc. Lond. B 249:299–302.

Taylor, P. D., T. Day, and G. Wild. 2007. Evolution of cooperation in a finite homogeneous graph. Nature 447:469.

Traulsen, A., and M. A. Nowak. 2006. Evolution of cooperation by multilevel selection. Proceedings of the National Academy of Sciences 103:10952–10955.

Trivers, R. L., and H. Hare. 1976. Haploidploidy and the evolution of the social insect. Science 191:249–263.

Van Baalen, M., and D. A. Rand. 1998. The unit of selection in viscous populations and the evolution of altruism. J. Theoret. Biol. 193:631–648.

Van Cleve, J. 2015. Social evolution and genetic interactions in the short and long term. Theoret. Popul. Biol. 103:2–26.

Van Cleve, J., and L. Lehmann. 2013. Stochastic stability and the evolution of coordination in spatially structured populations. Theoret. Popul. Biol. 89:75–87.

Wenseleers, T., and F. L. Ratnieks. 2004. Tragedy of the commons in Melipona bees. Proceedings of the Royal Society of London B: Biological Sciences 271(Suppl 5):S310–S312.

Wenseleers, T., and F. L. Ratnieks. 2006a. Comparative analysis of worker reproduction and policing in eusocial Hymenoptera supports relatedness theory. Am. Nat. 168:E163–E179.

Wenseleers, T., and F. L. Ratnieks. 2006b. Enforced altruism in insect societies. Nature 444:50.

Wenseleers, T., A. G. Hart, and F. L. Ratnieks. 2004a. When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. Am. Nat. 164:E154–E167.

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

Wenseleers, T., F. L. Ratnieks, and J. Billen. 2003. Caste fate conflict in swarm-founding social Hymenoptera: an inclusive fitness analysis. J. Evol. Biol. 16:647–658.

West, S. A., and A. Buckling. 2003. Cooperation, virulence and siderophore production in bacterial parasites. Proc. R Soc. Lond. B Biol. Sci. 270:37–44.

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

West, S. A., A. S. Griffin, and A. Gardner. 2007b. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. J. Evol. Biol. 20:415–432.

West, S. A., E. T. Kiers, E. L. Simms, and R. F. Denison. 2002. Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proc. R Soc. Lond. B Biol. Sci. 269:685–694.

West, S. A., M. G. Murray, C. A. Machado, A. S. Griffin, and E. A. Herre. 2001. Testing Hamilton's rule with competition between relatives. Nature 409:510.

West, S. 2009. Sex allocation. Princeton Univ. Press, Princeton.

Westneat, D., and C. W. Fox (Eds.). 2010. Evolutionary behavioral ecology. Oxford Univ. Press, Oxford.

Wild, G., and S. A. West. 2007. A sex allocation theory for vertebrates: combining local resource competition and condition-dependent allocation. Am. Nat. 170:E112–E128.

Wild, G., A. Gardner, and S. A. West. 2009. Adaptation and the evolution of parasite virulence in a connected world. Nature 459:983.

Wild, G. 2011. Direct fitness for dynamic kin selection. J. Evol. Biol. 24:1598–1610.

Yeh, A. Y. C., and A. Gardner. 2012. A general ploidy model for the evolution of helping in viscous populations. J. Theoret. Biol. 304:297–303.

The references listed here are for Chapters 1 and 9 only, as Chapters 2-8 each contain their own reference lists.

Bibliography

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A., Andersson, M., Andre, J.-B., Van Baalen, M., Balloux, F., Balshine, S., et al. (2011). Inclusive fitness theory and eusociality. *Nature*, 471(7339):E1.

Allen, B. (2015). Inclusive fitness theory becomes an end in itself.

Allen, B. and Nowak, M. A. (2015). Games among relatives revisited. *Journal of theoretical biology*, 378:103–116.

Allen, B. and Nowak, M. A. (2016). There is no inclusive fitness at the level of the individual. *Current Opinion in Behavioral Sciences*, 12:122–128.

Allen, B., Nowak, M. A., and Wilson, E. O. (2013). Limitations of inclusive fitness. *Proceedings of the National Academy of Sciences*, 110(50):20135–20139.

Bebbington, K. and Kingma, S. A. (2017). No evidence that kin selection increases the honesty of begging signals in birds. *Evolution letters*, 1(3):132–137.

Bianconi, G., Zhao, K., Chen, I. A., and Nowak, M. A. (2013). Selection for replicases in protocells. *PLoS Comput Biol*, 9(5):e1003051.

Boerlijst, M. C. and Hogeweg, P. (1991). Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites. *Physica D: Nonlinear Phenomena*, 48(1):17–28.

Boerlijst, M. C. and Hogeweg, P. (1995). Spatial gradients enhance persistence of hypercycles. *Physica D: Nonlinear Phenomena*, 88(1):29–39.

Boomsma, J. J. and Gawne, R. (2018). Superorganismality and caste differentiation as points of no return: how the major evolutionary transitions were lost in translation. *Biological Reviews*, 93(1):28–54.

Bourke, A. F. (2011a). *Principles of social evolution*. Oxford University Press.

Bourke, A. F. (2011b). The validity and value of inclusive fitness theory. *Proceedings of the Royal Society B: Biological Sciences*, 278(1723):3313–3320.

Caro, S. M., West, S. A., and Griffin, A. S. (2016). Sibling conflict and dishonest signaling in birds. *Proceedings of the National Academy of Sciences*, 113(48):13803–13808.

Cavalli-Sforza, L. L. and Feldman, M. W. (1978). Darwinian selection and "altruism". *Theoretical population biology*, 14(2):268–280.

Cooper, G. A. and West, S. A. (2018). Division of labour and the evolution of extreme specialization. *Nature ecology & evolution*, 2(7):1161.

Cronhjort, M. B. and Blomberg, C. (1997). Cluster compartmentalization may provide resistance to parasites for catalytic networks. *Physica D: Nonlinear Phenomena*, 101(3-4):289–298.

Darwin, C. (1859). *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. John Murray.

Davies, N., Krebs, J., and West, S. (2012). *An introduction to behavioural ecology. 4th edition*. Oxford: Wiley-Blackwell.

Des Marais, D. J., Nuth III, J. A., Allamandola, L. J., Boss, A. P., Farmer, J. D., Hoehler, T. M., Jakosky, B. M., Meadows, V. S., Pohorille, A., Runnegar, B., et al. (2008). The nasa astrobiology roadmap. *Astrobiology*, 8(4):715–730.

Fisher, R., Bell, T., and West, S. (2016). Multicellular group formation in response to predators in the alga chlorella vulgaris. *Journal of evolutionary biology*, 29(3):551–559.

Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford University Press.

Fisher, R. M., Henry, L. M., Cornwallis, C. K., Kiers, E. T., and West, S. A. (2017). The evolution of host-symbiont dependence. *Nature communications*, 8:15973.

Foster, K. (2009). A defense of sociobiology. In *Cold Spring Harbor symposia on quantitative biology*, volume 74, pages 403–418. Cold Spring Harbor Laboratory Press.

Foster, K. R. and Wenseleers, T. (2006). A general model for the evolution of mutualisms. *Journal of evolutionary biology*, 19(4):1283–1293.

Frank, S. A. (1994). Kin selection and virulence in the evolution of protocells and parasites. *Proceedings of the Royal Society of London B: Biological Sciences*, 258(1352):153–161.

Frank, S. A. (1998). *Foundations of social evolution*. Princeton University Press.

Frank, S. A. (2012a). Natural selection. iv. the price equation. *Journal of evolutionary biology*, 25(6):1002–1019.

Frank, S. A. (2012b). Natural selection. v. how to read the fundamental equations of evolutionary change in terms of information theory. *Journal of evolutionary biology*, 25(12):2377–2396.

Frank, S. A. (2013). Natural selection. vii. history and interpretation of kin selection theory. *Journal of Evolutionary Biology*, 26(6):1151–1184.

Gardner, A. (2017). The purpose of adaptation. *Interface focus*, 7(5):20170005.

Gardner, A. and Grafen, A. (2009). Capturing the superorganism: a formal theory of group adaptation. *Journal of evolutionary biology*, 22(4):659–671.

Gardner, A., West, S. A., and Wild, G. (2011). The genetical theory of kin selection. *Journal of evolutionary biology*, 24(5):1020–1043.

Grafen, A. (1982). How not to measure inclusive fitness. *Nature*, 298(5873):425.

Grafen, A. (1984). Natural selection, kin selection and group selection. *Behavioural ecology: An evolutionary approach*, 2:62–84.

Grafen, A. (1985). A geometric view of relatedness. *Oxford surveys in evolutionary biology*, 2(2).

Grafen, A. (2006). Optimization of inclusive fitness. *Journal of Theoretical Biology*, 238(3):541–563.

Grafen, A. (2014). The formal darwinism project in outline. *Biology & Philosophy*, 29(2):155–174.

Grafen, A. (2015). Biological fitness and the fundamental theorem of natural selection. *The American Naturalist*, 186(1):1–14.

Hamilton, W. D. (1964). The genetical theory of social behavior. i and ii. *Journal of Theoretical Biology*, 7(1):1–52.

Hamilton, W. D. (1970). Selfish and spiteful behaviour in an evolutionary model. *Nature*, 228(5277):1218–1220.

Higgs, P. G. and Lehman, N. (2015). The rna world: molecular cooperation at the origins of life. *Nature Reviews Genetics*, 16(1):7–17.

Horneck, G., Walter, N., Westall, F., Grenfell, J. L., Martin, W. F., Gomez, F., Leuko, S., Lee, N., Onofri, S., Tsiganis, K., et al. (2016). Astromap european astrobiology roadmap. *Astrobiology*, 16(3):201–243.

Kapsetaki, S. E., Fisher, R. M., and West, S. A. (2016). Predation and the formation of multicellular groups in algae. *Evolutionary Ecology Research*, 17(5):651–669.

Karlin, S. and Matessi, C. (1983). The eleventh ra fisher memorial lecture-kin selection and altruism. *Proceedings of the Royal society of London. Series B. Biological sciences*, 219(1216):327–353.

Koschwanez, J. H., Foster, K. R., and Murray, A. W. (2013). Improved use of a public good selects for the evolution of undifferentiated multicellularity. *Elife*, 2:e00367.

Krebs, J. R. and Davies, N. B. (1978). *Behavioural ecology: an evolutionary approach*. John Wiley & Sons.

Krebs, J. R. and Davies, N. B. (1987). *An introduction to behavioral ecology. 2nd edition*. John Wiley & Sons.

Krebs, J. R. and Davies, N. B. (2009). *Behavioural ecology: an evolutionary approach. 3rd edition*. John Wiley & Sons.

Lehmann, L., Alger, I., and Weibull, J. (2015). Does evolution lead to maximizing behavior? *Evolution*, 69(7):1858–1873.

Lehmann, L., Mullon, C., Akcay, E., and Van Cleve, J. (2016). Invasion fitness, inclusive fitness, and reproductive numbers in heterogeneous populations. *Evolution*, 70(8):1689–1702.

Lehmann, L. and Rousset, F. (2010). How life history and demography promote or inhibit the evolution of helping behaviours. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1553):2599–2617.

Levin, S. and West, S. (2017a). Kin selection in the rna world. *Life*, 7(4):53.

Levin, S. R., Scott, T. W., Cooper, H. S., and West, S. A. (2017). Darwin's aliens. *International Journal of Astrobiology*, 18(1):1–9.

Levin, S. R. and West, S. A. (2017b). The evolution of cooperation in simple molecular replicators. *Proceedings of the Royal Society B: Biological Sciences*, 284(1864):20171967.

McCaskill, J. S., Füchslin, R. M., and Altmeyer, S. (2001). The stochastic evolution of catalysts in spatially resolved molecular systems. *Biological chemistry*, 382(9):1343–1363.

Nowak, M. A. and Allen, B. (2015). Inclusive fitness theorizing invokes phenomena that are not relevant for the evolution of eusociality. *PLoS biology*, 13(4):e1002134.

Nowak, M. A., McAvoy, A., Allen, B., and Wilson, E. O. (2017). The general form of hamilton's rule makes no predictions and cannot be tested empirically. *Proceedings of the National Academy of Sciences*, pages 5665–5670.

Nowak, M. A., Tarnita, C. E., and Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466(7310):1057–1062.

Okasha, S. and Martens, J. (2016). Hamilton's rule, inclusive fitness maximization, and the goal of individual behaviour in symmetric two-player games. *Journal of evolutionary biology*, 29(3):473–482.

Pepper, J. W. (2000). Relatedness in trait group models of social evolution. *Journal of Theoretical Biology*, 206(3):355–368.

Price, G. R. (1972). Fisher's 'fundamental theorem' made clear. *Annals of human genetics*, 36(2):129–140.

Queller, D. C. (1985). Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature*, 318(6044):366–367.

Queller, D. C. (1992). A general model for kin selection. *Evolution*, 46(2):376–380.

Queller, D. C. (1996). The measurement and meaning of inclusive fitness. *Animal Behaviour*, 1(51):229–232.

Queller, D. C. (1997). Cooperators since life began. *The Quarterly Review of Biology*, 72(2):184–188.

Queller, D. C. (2016). Kin selection and its discontents. *Philosophy of Science*, 83(5):861–872.

Queller, D. C. (2017). Fundamental theorems of evolution. *The American Naturalist*, 189(4):345–353.

Queller, D. C. and Strassmann, J. E. (1998). Kin selection and social insects. *Bioscience*, 48(3):165–175.

Rousset, F. (2015). Regression, least squares, and the general version of inclusive fitness. *Evolution*, 69(11):2963–2970.

Sardanyés, J. and Solé, R. V. (2007). Spatio-temporal dynamics in simple asymmetric hypercycles under weak parasitic coupling. *Physica D: Nonlinear Phenomena*, 231(2):116–129.

Shay, J. A., Huynh, C., and Higgs, P. G. (2015). The origin and spread of a cooperative replicase in a prebiotic chemical system. *Journal of theoretical biology*, 364:249–259.

Smith, J. M. and Szathmáry, E. (1995). *The major transitions in evolution*. Oxford University Press.

Strassmann, J. E. and Queller, D. C. (2010). The social organism: congresses, parties, and committees. *Evolution: International Journal of Organic Evolution*, 64(3):605–616.

Szabó, P., Scheuring, I., Czárán, T., and Szathmáry, E. (2002). In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature*, 420(6913):340–343.

Taylor, P. (2017). Inclusive fitness in finite populations – effects of heterogeneity and synergy. *Evolution*, 71(3):508–525.

Taylor, P. D. (1992). Altruism in viscous populations—an inclusive fitness model. *Evolutionary ecology*, 6(4):352–356.

Uyenoyama, M. K. and Feldman, M. (1982). Population genetic theory of kin selection. ii. the multiplicative model. *The American Naturalist*, 120(5):614–627.

West, S. A., Fisher, R. M., Gardner, A., and Kiers, E. T. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, 112(33):10112–10119.

West, S. A. and Gardner, A. (2013). Adaptation and inclusive fitness. *Current Biology*, 23(13):R577–R584.

Westneat, D. and Fox, C. W. (2010). *Evolutionary behavioral ecology*. Oxford University Press.