UNIVERSITY OF SOUTHAMPTON

The Evolution of Biological Individuality by Social Niche Construction

by

Paul Ryan

Thesis for the degree of Doctor of Philosophy

in the
Faculty of Physical Sciences and Engineering
School of Electronics and Computer Science

March 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
SCHOOL OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

by Paul Ryan

The biological world as we see it today has a part-whole hierarchical structure. For example, eusocial societies are made up of many organisms, multicellular organisms are made up of many cells, those cells contain numerous organelles and so on. This hierarchical organisation is thought to have evolved over a long period of time in a series of events known as 'evolutionary transitions in individuality'. Evolutionary transitions present an interesting challenge for evolutionary theory because they involve changes in the hierarchical level at which the evolutionary process itself acts. This thesis is intended as a contribution to theoretical work aiming to explain such transitions in the hierarchical structure of life.

Evolutionary transitions are extreme cases of the evolution of cooperation. Social evolution theory is the part of evolutionary theory that tries to explain the evolution of cooperation. It typically takes an externalist explanatory stance, explaining cooperative behaviour in terms of external factors (e.g. genetic relatedness) that make cooperation sustainable. In this thesis, I move from an externalist to an interactionist explanatory stance, in the spirit of Lewontin and the niche construction theorists. I develop the theory of social niche construction, which has it that biological entities are both the subject and object of their own social evolution. That is, the niche in which social behaviour occurs is not entirely externally defined but is partly modified by the organisms in it. Then, cooperation and the social niche modifier traits supporting it can each evolve as evolutionary responses to the other. This claim is supported by detailed argument and by simulation modelling.

Some important social niche modifiers enabling cooperation (e.g. life-history bottlenecks) have the side-effect of raising the hierarchical level at which the evolutionary process acts. This is because modifier traits acting to align the fitness interests of lower-level units (e.g. cells) in a collective also diminish the extent to which those units are bearers of heritable fitness variance, while augmenting the extent to which collectives of such units (e.g. multicellular organisms) are bearers of heritable fitness variance. So while there is no *selection-for* evolutionary transitions in individuality, there is *selection-of* the sufficient conditions for transitions to occur. My explanation for evolutionary transitions is couched only in terms of evolutionary self-interest of the lower-level units, so avoiding many of the problems that befall alternative accounts.

Declaration of Authorship

I, Paul Ryan, declare that the work presented in this thesis is my own and has been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

Signed:

Date:

Supervisors

- Primary supervisor: Richard A. Watson
- Secondary supervisor: C. Patrick Doncaster

Acknowledgements

Many thanks are due to my PhD supervisor Richard Watson for many hours of to-and and fro-ing over the ideas contained in this thesis. Thanks also to Andrew Bourke, Markus Brede, Ellen Clarke, Patrick Doncaster, Andy Gardner, Adam Jackson, Guy Jacobs, Rick Michod, Silvia De Monte, John Odling-Smee, Will Ratcliff, Simon Powers, Paul Rainey, Carl Simpson and Simon Tudge for useful and stimulating discussions.

I am grateful to the Institute for Complex Systems Simulation for giving me the opportunity to pursue this research project, supported by EPSRC Doctoral Training Centre grant (EP/G03690X/1). I acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

This thesis is dedicated to Claire, for all her love, patience and support.

Contents

List of Figures

Chapter 1

Introduction

1.1 Overview of thesis

This thesis is about evolutionary transitions in individuality (also known as 'major transitions in evolution' after Maynard Smith and Szathmáry, 1995).

The biological world as we see it today has a part-whole hierarchical structure. (For example, eusocial societies are made up of many organisms, organisms are made up of many cells, those cells contain numerous organelles and so on.) It has long been recognised that entities at several levels in that hierarchy could, in principle, be evolutionary units in virtue of potentially having heritable variation in fitness (Lewontin, 1970). The well-known debate about the relative strength and practical significance of the evolutionary process at each of these levels (e.g. Williams, 1966; Wilson, D.S., 1975; Dawkins, 1976; Vrba, 1984) does not address the origin of the hierarchy itself (Griesemer, 2000).

Explaining the evolution of new hierarchical levels at which the evolutionary process acts is the business of the Major Evolutionary Transitions research programme (Margulis, 1981; Buss, 1987; Maynard Smith, 1988; Maynard Smith and Szathmáry, 1995; Michod, 1999). Here, the thing to be explained is not whether this or that level is the 'real' or 'main' level of selection but how the levels arose in the first place. Okasha (2005b) describes the earlier levels-of-selection question as formulated in a 'synchronic' way, while the Major Transitions approach is to formulate the question in a 'diachronic' way. This approach gets its name as follows. The hierarchical organisation of life is thought to have evolved over a long period of time in a series of events known as 'major transitions in evolution' or 'evolutionary transitions in individuality' (ETIs - I will use the latter term in this thesis). These transitions involve a coalescing or coming-together of entities over evolutionary time, yielding new levels of biological individuality (Queller, 1997). After a transition, entities that could previously survive and reproduce in their own right can only do so as part of a larger, collective whole (Maynard Smith and Szathmáry, 1995). The challenge is to explain these transitions in Darwinian terms.

This thesis aspires to be a contribution to the growing body of theoretical work on evolutionary transitions in individuality. The achievable goals are more modest. The thesis aims to persuade the reader that a few things are wrong with current thinking on evolutionary transitions and it offers some suggestions on how they might be improved.

I present an outline explanation for how some evolutionary transitions might occur, involving a process called social niche construction. This is supported with verbal argument and demonstrated with simulation modelling. There is already another well-known theory that attempts to explain the same - the Fitness Export Theory (also known as Fitness Decoupling Theory) associated with Michod and co-authors. I show Fitness Export Theory to be problematic in a number of ways. This critique leads me to argue that some of the conceptual assumptions of recent theoretical work on evolutionary transitions (particularly the role of the concept of fitness) are not sound. I propose some repairs and suggest a way forward.

Major evolutionary transitions present a multi-faceted challenge for evolutionary theory. The distinction between philosophical and theoretical contributions is not sharp and that is reflected in the inter-disciplinary character of this thesis.

There is a glossary of terms in an appendix just before the references.

1.2 Claims of the thesis

For transparency and in the interests of getting straight to the point, I set out the claims of the thesis here. To avoid any ambiguity, I have set out a clear definition of what I take to be an evolutionary transition in individuality in Section 1.4. The claims will be made more fully and justified in the chapters that follow. In this thesis, I make the following claims:

- Some evolutionary transitions in individuality (ETIs), as defined in Section 1.4 below, can be explained by social niche construction.
- Social niche construction employs an interactionist explanatory stance. I claim that the interactionist explanatory stance matches the causal structure of the world better than the externalist explanatory stance, when the thing to be explained is evolutionary transitions in individuality. (The externalist explanatory stance is usually adopted in theoretical work on the evolution of cooperation and evolutionary transitions).
- Social niche construction provides an evolutionary explanation that proceeds without assuming the prior existence of any higher-level evolutionary process, even though the product of such transitions can be a higher-level evolutionary process.
- The onset of an evolutionary transition in individuality is characterised by increasing alignment of fitness interests among pre-existing lower-level units, such that pre-existing lower-level units gain inclusive fitness from the transition (other things equal). Therefore,

the onset of an evolutionary transition can be explained in terms of pre-existing lower-level units pursuing their own inclusive fitness interests. (The significance of this is that I argue against the widely accepted position that says evolutionary transitions involve altruistic self-sacrifice on the part of the lower-level units.)

- The currently best-accepted theory for explaining evolutionary transitions in individuality (i.e. Fitness Export Theory) is unworkable and does not explain what it purports to explain.
- In the context of evolutionary transitions in individuality, fitness is best thought of a single quantity wherever it is found. I claim that adopting ontological monism about fitness solves some of the problems I find with Fitness Export Theory and with biological individuality more generally (and is consistent with my own theoretical account). That is, I reject the widely held assumption that fitness involves counting offspring (and I reject variants based on reproduction, based on counting things and based on relative measures).

1.3 Disambiguating units of selection and units of evolution

Before I state the explanatory target of this thesis in the next section, it will be necessary to gain greater precision about what I mean by the terms 'unit of evolution' and 'unit of selection'. A good deal of confusion has been introduced by a failure to disambiguate these concepts.

As Lewontin (1970) famously pointed out, Darwin's theory is neutral about the substrate on which the evolutionary process operates. Whenever a population of entities (whatever they might be) engage in reproduction with heredity and whenever their reproductive success is a function of their heritable characters and those characters vary, then evolutionary change will ensue. In a hierarchical setting, it is then possible that entities at more than one hierarchical level could possess heritable fitness variance at the same time. The level or levels of the hierarchy at which heritable fitness variance obtains is/are often referred to as the 'level of selection' but there has been and remains serious confusion surrounding the use of this term.

'Level of selection' is often used in contexts where it clearly means 'level at which there is heritable variation in fitness' (e.g. the titles of Lewontin, 1970; Keller, 1999; Okasha, 2006) but taken literally 'level of selection' does not say anything about heritability, only selection. Consider the following example of why this matters. In MLS1 trait-group models[1] (e.g. Wilson, D.S., 1975; Powers et al., 2011), particles and collectives are both units of selection in the strict sense (because both have differential fitness as a function of differential character). But in such models, only particles have heritability - collectives do not. As Maynard Smith (1987, p.121) pointed out, collectives in such models are units of selection in a strict sense (because their fitnesses covary with their characters) but they are *not* units of evolution (because there is no

[1] See Section 2.4 on page 24 for a summary of this well known class of model, that features prominently in the literature on group selection.

collective-level heritability in such models, by design). Following Maynard Smith, I prefer the term 'unit of evolution' over 'unit of selection' when I mean a unit with heritable variation in fitness.

Covariance equations due to Price (see Price, 1970; Gardner, 2008) can be used to make the unit of selection / unit of evolution distinction more precise. A hierarchical level with entities having fitness w and character z is a level at which we find **units of selection** when:

$$cov(w, z) \neq 0 \tag{1.3.1}$$

A hierarchical level with entities having fitness w, character z and mean offspring character z' is a level at which we find **units of evolution** when:

$$cov(w, z') \neq 0 \tag{1.3.2}$$

This second equation can be understood as breaking down into two parts:

$cov(w, z) \neq 0$	(selection: fitness varies with character)
$cov(z, z') \neq 0$	(heredity: offspring character resembles parent character)

(As Okasha (2006, p.37) points out, the equivalence between the single-part version of the Price formalism for units of evolution (Equation 1.3.2) and the two-part version above is not fully general because covariance is not generally a transitive relation. This two-part Price formalism should be understood as assuming that the two covariances are strong enough that there *is* covariance between entity fitness and offspring character, as in Equation 1.3.2.)

In Chapter 2 I will survey the literature on a (relatively) novel approach to biological individuality that sees biological individuality closely bound up with the concept of a 'unit of evolution'. This approach can plausibly overcome many of the problems giving an account biological individuality and is the approach I adopt in this thesis.

1.4 Definition of the thing to be explained in the thesis

There are potentially multiple understandings of the term 'evolutionary transition in individuality'[2]. This section aims at clarity on exactly what I am trying to explain in this thesis.

I assume that there is something in general that can be said about evolutionary transitions (West et al., 2015). That is, despite their involving radically different types of things, vastly different spatial scales and being separated by geologically long periods of time, I assume there is something in common among (at least some) evolutionary transitions about which it is possible

[2]For example, Maynard Smith and Szathmáry (1995) consider evolutionary transitions to involve changes in the way information is transmitted from one generation to the next. I do not explore that possibility here.

to generalise. This is itself a potentially controversial claim (Calcott, 2011) but its defence is beyond the scope of this thesis.

To keep things tractable, I have adopted a narrow definition of evolutionary transitions in individuality. Throughout this thesis, including this chapter, I am interested in how the evolutionary process can come to operate at multiple hierarchical levels, from an initial state where it operates only at one level. For simplicity, I use only two levels in my analyses (and the new level is assumed to be above rather than below the pre-existing level of individuality). I assume the existence of some sort of grouping; the origin of this population structure is not the target of explanation in this thesis.

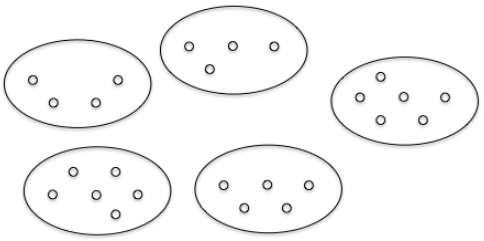

FIGURE 1.1: A two-level population structure. Particles are nested inside collectives. The two levels are strictly relative - particles at one level might be collectives at a lower level (that is not shown).

Definition: A population consists of a number of particles, nested within collectives. Particles have characters (properties) that may vary. Particles have fitness (a concept to be expanded upon). Collectives have characters (properties) that may vary. Collectives may have fitness (a concept to be expanded upon). These terms (following Okasha, 2006) are used throughout the thesis, including this chapter - hence their definition here - to discuss both abstract hierarchical scenarios and hierarchical examples from the biological world, e.g. particles/collectives, cells/multicellular organisms, ants/eusocial colonies or zooids/siphonophores.

I take a character-fitness covariance approach (Price, 1970; Okasha, 2006; Clarke, 2010) to evolutionary individuality[3]. Evolutionary change is expected among any population of reproducing entities if their characters covary with their fitnesses and their characters are heritable. The model of an evolutionary transition in individuality (ETI) that I take to be my explanandum is as follows:

- Assume a two-level hierarchical population structure, featuring particles nested in collectives (as in Figure 1.1).
- Before an evolutionary transition in individuality, particle fitness covaries with mean particle offspring character (not so of collectives)

[3]This approach will be expanded upon at length in subsequent chapters. For now I use 'character-fitness covariance approach' only as a label to refer to a way of thinking about evolutionary change and without any commitment to what the characters and fitnesses are predicated of and how they are measured.

- After an evolutionary transition in individuality, collective fitness covaries with mean collective offspring character (not so of particles, or only to a limited extent)

The change from 'before' to 'after', summarised in Table 1.1, is the thing I take myself to be trying to explain in this thesis[4].

BEFORE ETI	particle-level	collective-level
selection	$cov(w, z) \approx$ high	$cov(W, Z) \approx 0$
heredity	$cov(z, z') \approx$ high	$cov(Z, Z') \approx 0$
AFTER ETI	particle-level	collective-level
selection	$cov(w, z) \approx 0$	$cov(W, Z) \approx$ high
heredity	$cov(z, z') \approx$ still high	$cov(Z, Z') \approx$ high

TABLE 1.1: Rough character-fitness-covariance characterisation of the start and end states of an evolutionary transition in individuality, as defined in this thesis. Particle level fitness and character are w and z respectively. Collective level fitness (a concept to be expanded upon below) and character are W and Z respectively. Absolute values are implied ('high' can be positive or negative). The thing to be explained is how an evolutionary process initially operating only at particle level can cause a shift from the 'before' to 'after' state.

The challenge is to explain how an evolutionary process could bring about such a transition. The difficulty is that it is hard to see how an evolutionary process at particle level could effect an upward transition in the hierarchical level at which that very process operates. How could the evolutionary process at collective-level ever get started, given that it seems to be faced with a bootstrapping problem?

1.5 What is out of scope

Note that the character-fitness covariance approach I adopt is somewhat different to talk of post-transitions particles 'only being capable of reproduction as part of a larger whole' (after Maynard Smith and Szathmáry (1995)) that often appears in more casual descriptions of evolutionary transitions. Some authors understand evolutionary transitions in terms of the evolution of collective-level reproduction mechanisms (e.g. Griesemer (2000), Rainey and Kerr (2011); Carl Simpson, pers. comm., 'The evolution of reproduction during evolutionary transitions in individuality'). I do not foreground the evolution of new reproductive mechanisms in this thesis, although they are implicated in what I have to say in Chapters 3 and 4. Indeed, I take it that any complex collective-level adaptation could only evolve once an evolutionary process is acting sufficiently strongly at collective level, so the subject of this thesis (the evolutionary origin of collective-level evolution) - is logically and temporally prior to any narrative involving the evolution of such mechanisms.

[4]See Clarke (2014, Section 2) for a similar characterization of the thing to be explained about evolutionary transitions.

The character-fitness covariance approach was chosen because its common-sense breakdown of evolutionary change into within-group and between-group components is widely used and intuitively straightforward. Character-fitness covariance is not the only approach one might take to analysing multi-level selection. Alternative approaches have been used (e.g. contextual analysis - Heisler and Damuth (1987); Goodnight et al. (1992)) but are outside the scope of this thesis.

One might question whether a character-fitness covariance at a level might be misleading. Covariance and correlation do not imply causation. A character-fitness covariance might be spurious (Okasha 2015; Krupp, forthcoming). It also might be due to a causally significant character-fitness covariance at another hierarchical level - a cross-level byproduct (Sober, 1984; Okasha, 2006). Metaphysical worries about correlation and covariance, while well-founded, are beyond the scope of this thesis. In many of the cases considered in this thesis, these worries are sidelined because selection is assumed (or defined) to act in different directions at each level, so outcomes are not causally overdetermined. Furthermore, I do not consider any definitions of collective-level selection based on a requirement that the collective characters under selection bear an emergent rather than aggregative relation with particle characters. That is, the so-called 'emergent character requirement' (Vrba, 1989; Okasha, 2006, p.106-107) is out of scope.

1.6 Chapter-by-chapter summary

Chapter 2 is a focused literature review[5], concentrating on the long-standing problem of defining biological individuality and especially on one particular response to it, which I call the 'evolutionary-unit-theory of biological individuality' (e.g. Buss, 1987; Godfrey-Smith, 2009; Clarke, 2010; Rainey and Kerr, 2011; Clarke, 2014). The problem of biological individuality (as I set it up here) is that it is difficult to define a principled way of distinguishing between biological individuals, their parts and groups of them. The problem has been around since ancient times but a relatively new approach to it has been making good progress recently - the 'evolutionary-unit-theory of biological individuality'. On that theory, biological individuals are (roughly) whatever things are evolutionary units - bearers of heritable fitness variance - in a given situation. This understanding of individuality can potentially apply to multiple hierarchical levels at the same time (Buss, 1987; Okasha, 2006; Godfrey-Smith, 2009) but there is a hard tradeoff that means more individuality at one level means less at others (Clarke, 2010; Godfrey-Smith, 2009, Chapter 5). Many past attempts to define biological individuality have latched onto the existence of certain individuating mechanisms that seem to underpin particular cases of biological individuality (often with an animal focus). These mechanisms can be quite diverse but old favourites include single-celled life-history bottlenecks (Dawkins, 1976) and segregated germ lines (Weismann, 1903; Maynard Smith, 1958) (see Table 3.1 in Chapter 3). The evolutionary-unit-theory of biological individuality says these individuating mechanisms have the function of causing heritable fitness variance to dominate at the hierarchical level at which

[5]There is plenty more reviewing of literature throughout the thesis at appropriate places.

we recognise individuality, relative to other levels. That is, individuation mechanisms are seen as contingent ways to *instantiate* individuality at a level but are not themselves essential to individuality (Clarke, 2010). In this thesis, I am interested in the *evolution* of those individuating mechanisms. That is, I am interested in the process of which they are the product.

Chapter 3 on social niche construction is the theoretical core of the thesis. The term 'social niche construction' comes to me from Powers (2010; 2011) and gets its name as follows. Social evolution theory tries to explain the evolution of social behaviours including altruism and cooperation (Hamilton, 1964; Wilson, E.O., 1975; Bourke, 2011); niche construction is any process in which organisms modify their own environment in such a way as to influence the conditions of their own evolution (Odling-Smee et al., 2003; Laland and Sterelny, 2006). The term 'social niche construction' denotes the application of niche construction theory to social evolution.

The easiest way to start with social niche construction is to consider a single-level population, leaving hierarchical scenarios and transitions for later. Biological cooperation was a problem for Darwin - the 'one special difficulty' for his theory. Social evolution theory typically tackles this problem by taking an externalist explanatory stance, explaining cooperative behaviour by pointing to additional factors (e.g. relatedness or vertical transmission) that make cooperation sustainable. Those additional factors remain exogenous to the explanation. For example, reproductive altruism among the worker caste in honey bees is initially puzzling from an evolutionary perspective - how could such a trait evolve and persist when it seems so obviously doomed? Social evolution theory might offer the explanation that workers are more related to their sisters (some of whom may go on to found new colonies) than they would be to their own offspring. In such a case, the prevailing relatedness structure would be taken to be something external to the bees and that is what makes this an example of an externalist explanation. Social niche construction is different in that it takes an *interactionist explanatory stance*. The idea is that biological entities may have some degree of control, however small, over the circumstances of their own social evolution. (To continue with the relatedness example, they may have some degree of control over the genetic relatedness they share with those with whom they interact). So it is possible for cooperation and the circumstances supporting it each to evolve as evolutionary responses to the other. I call the circumstances determining the direction and strength of selection on social behaviour (e.g. cooperation) a 'social niche' and any traits organisms may have to change their own social niche I call 'social niche modifiers'. I provide lots of examples of social niche modifiers in Table 3.1, including fair meiosis, uniparental inheritance of mtDNA, kin-recognition mechanisms in colonial ascidians, mycelial fungi and cellular slime moulds, cell-cycle synchronization mechanisms in myxomycetes and mechanisms of policing, punishment and coercion in eusocial insect societies. (To further continue the relatedness example - cooperation among honey bees was a 'special difficulty' for Darwin because he made wrong assumptions about their social niche.)

Now consider a two-level population, with particles nested within collectives (as in Figure 1.1). Some social niche modifiers can be though of as doing two things:

1) They make circumstances more favourable for cooperation (and there is *selection-for* this[6]).
2) They have the side-effect of changing the hierarchical level at which the evolutionary process acts (and there is *selection-of* this).
The first point involves changing the effective game being played (the social niche) among particles, furthering the alignment of their fitness interests. The second point is more subtle. A social niche modifier can change the hierarchical level at which character-fitness covariance (and thus individuality) obtains by altering the way fitness variance is partitioned at each level and by altering the extent to which there is faithfulness of heredity for entities at each level. For example, life-history bottlenecks (a very common social niche modifier) have both properties.

The main claim of the chapter (and thesis) is that *social niche construction can explain upward transitions in the hierarchical level at which the Darwinian machine operates*. The explanation starts by invoking only a particle-level evolutionary process, even though the outcome of the process is a collective-level evolutionary process. I suggest that social niche construction based explanations for cooperation are better than externalist ones because they more closely match the causal structure of the world. And I suggest that social niche construction based explanations for evolutionary transitions in individuality are better than the currently best-accepted theory because they do not run into the same cart-before-horse logical and metaphysical problems (I return to this in Chapters 4 and 5).

The general theory of social niche construction is presented as verbal argument. However, the theoretical treatment of social niche construction among like-kinds (sometimes leading to fraternal evolutionary transitions) is examined in more technical detail. I explore the space of possible two-player two-strategy symmetric games (including but not limited to the well-known Prisoners Dilemma) and how these games are modified by the introduction of positive assortment of strategies. Many (but not all) social niche modifiers operate by introducing assortment.

Chapter 4 offers an in-depth example of the evolution of a life-history bottleneck by social niche construction. A life-history bottleneck is a single-celled stage (like a zygote or a spore) in the life cycle of a multi-cellular organism, or analogously a single-unit propagule in colonial organisms (like the single zooid that founds a new colony of marine bryozoans such as Flustra foliacea). Bottlenecks are very common in all parts of the tree of life and have evolved independently many times over. There are many hypotheses for their evolution and controversy remains on this issue. It has been noted by many that bottlenecks cause fitness variation within organisms/colonies (i.e. collectives) to be low, leaving fitness variation partitioned mainly between collectives. That is, bottlenecks are promoters of character-fitness covariance (and hence individuality) at collective level. It might be tempting to think bottlenecks evolved *because* they have this function but that would be to put the cart before the horse (Williams, 1992; Okasha, 2006). Bottlenecks may well act to maintain extant collective-level selection against subversion from below but that cannot be the reason they originally evolved. As Clarke (2014) points out, there is an apparent chicken-and-egg paradox here. Social niche construction theory side-steps it, providing

[6]There is selection-for a trait when the trait causes its bearers to be fitter than its non-bearers; there is selection-of a trait when the trait increases in frequency for some other reason (Sober, 1984, p.97).

an explanation that does not invoke collective-level selection to explain how collective-level selection gets started.

A simulation model is described. It employs a two-level hierarchically structured population, as in Figure 1.1. Particles have two traits: a Social Trait (Cooperate or Defect) and a Social Niche Modifying trait (being a trait that influences the number of particles used to found each offspring collective). Results are presented that show two things. Firstly, they show the concurrent evolution of cooperative social behaviour and ever-smaller bottleneck sizes (even starting from a state without small bottlenecks and with a base game that does not support cooperation). Secondly, the results employ the Price Approach (Clarke, 2010) to measure the relative strength of character-fitness covariance at each level in the two-level hierarchy and - importantly - how this changes during a simulation run. A diachronic change in the hierarchical level at which character-fitness covariance obtains is observed. The bottleneck model demonstrates that the hierarchical level at which character-fitness covariance obtains (informally 'the level of selection') can be shifted upward by an evolutionary process operating initially on only the lower-level entities (particles).

Scenarios that start off as conflicted social dilemmas (and without bottlenecking) can end harmoniously, with extreme bottlenecking and mutual cooperation widespread. The overall fitness of those particles is raised by such a transition, meaning that more of them can be sustained with a bottlenecked lifecycle than without it. Even though bottlenecking causes each collective to start life small, bottlenecked collectives grow larger than unbottlenecked ones. This is due to the fitness benefit yielded by the positive assortment of social behaviour brought about by bottlenecking.

The next two chapters argue for revising some basic concepts employed in theoretical work in evolutionary transitions in individuality. The first is critical of current theory, the second offers a constructive way forward. The inspiration for these chapters came directly from the modelling work.

Chapter 5 offers a critical examination of the currently best-accepted theory about how ETIs occur - the Fitness Export Theory (sometimes called Fitness Decoupling Theory) associated with Michod co-authors (Michod, 1999). The central intuition in Fitness Export Theory is one of self-sacrifice for the greater good, with fitness sacrifices being made by the particle level and the 'greater good' being the increased collective-level fitness so produced. This is explained in terms of the evolution of altruism among particles: altruism causes particles to decrease in fitness, due to paying the costs of altruism. At the same time, altruism causes collectives to increase in fitness, due to receiving the benefits of altruism. This transfer of fitness from particle level to collective level is what the phrase 'export of fitness' refers to. The transfer of fitness to the higher level amounts to a transfer of evolutionary individuality to the higher level. Once the evolutionary process has started to act at collective level, it is subsequently possible for mechanisms (e.g. life-history bottlenecks or germ-soma separation) to evolve that suppress within-collective conflict, so consolidating individuality at that level.

I am critical of this theory in several ways. I object to the self-sacrifice intuition, finding it unsupportable on any reasonable understanding of what evolutionary self-interest is. I offer a diagnosis of where Fitness Export Theory has gone wrong, finding that it runs together the concepts of fitness and individuality (fitness-*variance* being a necessary ingredient of the latter). I also discuss possible reasons for why it has seemed persuasive to many, despite its shortcomings.

The problems I raise with Fitness Export Theory lead naturally to questions about what fitness actually *is*. **Chapter 6** examines the concept of fitness as it features in theoretical work on evolutionary transitions in individuality. I find that a fitness concept borrowed from population biology has been uncritically applied to the domain (including Fitness Export Theory) but is not fit for purpose in that context. One problem is that it involves comparing counts of things, which wrongly assumes the things are comparable (and countable). Another problem is that this fitness is a relative concept and so cannot model hypothetical situations where gross changes in the absolute amount of living stuff is the driver of evolutionary change in which we are interested.

I argue for the adoption of a rather different understanding of fitness, inspired by (but not quite the same as) that proposed by Leigh Van Valen (1976; 1980; 1989). I call this VV-fitness[7]. It is a single absolute quantity, a measure of occupancy of the world though space and time, commensurable across all life, at all hierarchical levels and at all timescales. This understanding of fitness collapses down to the standard population biology view in the sense that heritable variation in it is still the driver of evolutionary change in a population. There are many advantages to such an understanding of fitness but I focus on those directly connected with evolutionary transitions in individuality. *Adopting such a conception of fitness allows us to make sense of what happens to fitness during evolutionary transitions.* It allows us to track the fitness interests of particles all the way through a transition, even at later stages where there is no within-collective fitness variance (and hence no within-collective evolution). In VV-fitness terms, a collective's fitness is identical with the inclusive fitness of its constituent particles (when measured over the same time scale, as it should be). The use of a well-known distinction between two understandings of collective-level fitness (fitness_1 and fitness_2), usually associated with two versions of multi-level selection (MLS1 and MLS2, Damuth and Heisler, 1988; Okasha, 2006), is critically examined and rejected. If one adopts a VV-fitness understanding, the driver of evolutionary transitions in individuality is unproblematically self-interest, with no paradoxes of self-sacrifice or altruism to be explained. While this approach makes a lot of sense (and is quite consistent with the lessons learned from the bottleneck model in Chapter 4 and with the natural world), it raises another question.

One might wonder how evolutionary theory could have progressed as far as it has with a flawed understanding of one of its central concepts. The answer I give is that for an evolutionist - someone interested in evolutionary change - it is not fitness itself that is of primary interest but

[7] Van Valen thought energy control was the basis of all fitness. I think he was probably right, but I do not need to include such a commitment into my fitness concept to get it to do the work I need done in this thesis. Hence VV-fitness is defined more conservatively than Van Valen's own fitness concept. Nevertheless, I retain his ontological monism about fitness - the fitness of apples and oranges *can* be compared - but not by counting things.

co-variance of fitness with character. For the class of organisms[8] upon which many foundational evolutionary models have been based (i.e. organisms that differ in fitness by differing in the number of equivalently-rated offspring they produce), it makes no difference if we multiply all fitnesses by a constant. This explains why more conventional fitness analyses get things right when they get it right. By adopting VV-fitness we can also understand what Fitness Export Theory gets right, how to repair what is wrong with it and how it could have been so plausible to many, despite its failings.

In the **final chapter** I revisit the list of Claims of the Thesis in Section 1.2 above. I consider each claim and show how it has been supported by the intervening chapters, with cross-references. The rest of the chapter is a discussion of a few themes emerging in the thesis.

1.7 List of publications and talks based on the thesis

Chapter 3 shares content with: Ryan, P.A., Powers, S.T. and Watson, R.A. (2016) Social Niche Construction and Evolutionary Transitions in Individuality, *Biology and Philosophy*, Volume 31, Number 1, pages 59-79.

I have presented material from this thesis at the following conferences:

- 'Philosophy of Biology in the UK', Christ's College, Cambridge, March 2014, Title: What happens to fitness during evolutionary transitions in individuality? (Chapter 5)
- 'Evolution of Multicellularity', Institució Catalana de Recerca i Estudis Avançats, Barcelona, September 2013, Title: The evolution of death-birth cycles in the life-history of multi-cellular organisms (Chapter 4)
- 'Cooperation and Major Evolutionary Transitions', University of California, Santa Barbara, February 2013, Title: The evolution of the life-history bottleneck (Chapter 4)
- 'London Evolutionary Research Network, Annual Conference', University College London, November 2012, Title: The evolution of the unicellular bottleneck in the life-history of multi-cellular organisms (Chapter 4)
- 'Evolution of Functional and Structural Complexity in Biology', King's College, Cambridge, September 2012, Title: The evolution of the unicellular bottleneck in the life-history of multi-cellular organisms (Chapter 4)
- 'Philosophy of Biology in the UK', All Souls College, Oxford, April 2012, Title: Bourke's concept of Social Group Transformation: a defence (Chapter 3)

[8] For reasons that will become clear below, I call these 'Weismannian organisms'.

Chapter 2

Literature review on the evolutionary-unit theory of biological individuality

2.1 Overview of chapter

The problem of biological individuality is that it is difficult to define or understand what makes something a biological individual (rather than a part of an individual, or a group of individuals) in a principled way. This chapter starts with a brief introduction to the problem of biological individuality, the various attempts that have been made to solve it since ancient times and why it matters. I then survey a relatively new response to the problem, arising out of work on major evolutionary transitions. I call this approach the 'evolutionary-unit theory of biological individuality'. It says that what is essential about biological individuals is that they are things that exhibit heritable variance in their fitness (and that many of the various prior attempts to explain individuality have each latched onto a contingent instantiator of this property). This approach is able to account for why many prior attempts to understand biological individuality seem at least partly right. It is also able to account for the gradient nature of individuality in many cases and for the way individuality changes over evolutionary time. This last point is the topic of this thesis, the evolution of biological individuality.

2.2 The problem of biological individuality

On the face of it, one might think that biological individuality is a straightforward and unproblematic concept. Indeed, it is so for humans, domestic animals and many other familiar

metazoan taxa. However, attempts to formulate a fully general theory of biological individuality have been problematic[1].

While the Major Transitions in Evolution research programme belongs largely to the twentieth century (and perhaps the late nineteenth), the problem of biological individuality is very much older and has a literature of its own. It is only since relatively recent times that these two literatures can be seen as converging. There is a long history of attempts to define biological individuality, motivated by a wide variety of considerations and informed by a wide variety of different pre-commitments. This activity has been ongoing since ancient times and remains controversial. The traditional problem of biological individuality is the failure to articulate a fully general theory of what individuality *is* - especially how to distinguish biological individuals from their parts and how to distinguish biological individuals from groups of individuals. (It is distinct from the problem of distinguishing living from non-living things.)

As with non-biological individuals, some have tried to define biological individuals in terms of spatial and temporal contiguity (Huxley, 1912; Hull, 1980; Anderson and McShea, 2001). There is a related cluster of individuality criteria relating to functional integration (Huxley, 1912; J.A. Wilson, 1999), physiological union (Anderson and McShea, 2001), metabolism and self-maintenance (Maturana and Varela, 1981). In a distinct approach, some have preferred to rely on genetic concepts, especially genetic homogeneity and uniqueness as definitive of biological individuals (Simpson, 1957; Janzen, 1977; Santelices, 1999). This is often bound up with another criterion - individuals as demarcated by life-history bottlenecks (Bonner, 1965; Dawkins, 1976). Genetic homogeneity and uniqueness can also be provided by a segregated germline (Weismann, 1903). The level of internal cooperation and conflict has been the favoured approach for some (Queller and Strassmann, 2009). Yet another approach has been to take the self/non-self distinction implemented by immune systems as definitive of individuals (Burnet, 1969; Pradeu, 2013). A further approach has been to identify individuals as the bearers of adaptations (Williams, 1966; Gardner, 2013; Huneman, 2013).

The list of potential candidate criteria for biological individuality is much longer than this selection (see Clarke, 2010, 2011b, for a thorough review) and the literature on them is very large indeed. For a sample of the confusion, see Santelices (1999), J.A. Wilson (1999), J.A. Wilson (2000), Ruiz-Mirazo et al. (2000), De Sousa (2005), Pepper and Herron (2008), Gardner and Grafen (2009), Godfrey-Smith (2009), Queller and Strassmann (2009), West and Kiers (2009), Dupré (2010), Folse and Roughgarden (2010), Folse (2011) and Herron et al. (2013).

Clarke and Okasha (2013) identify two distinct problems of biological individuality: the problem of vagueness and the problem of disagreement. The problem of vagueness is that, of each proposed definition of individuality, real-world cases are hard to categorise with certainty - the

[1] Does it matter? J.A. Wilson (1999) has argued that biology does not need to decide which things are individuals because no substantive issues depend upon it. While this may be true in some parts of biology, it is certainly not true from a methodological perspective where evolutionary biologists depend on counting things to measure fitness (Clarke, 2011b). In Chapter 6 I will question the assumption that fitness is always or mainly about counting things but that is not what is at stake here. The methodological point is that counting things really does go on a lot in evolutionary biology and, in those cases where it does, the counters need to know what to count.

Individuation criterion	**Zooid**	**Colony**
Reproduction	Yes	Yes
Life cycle	No	Yes
Genetics	No	Yes
Bottleneck life cycle	Yes	Yes
Germ soma separation	Yes	Yes
Spatial boundaries/contiguity	No	Yes
Histocompatibility	No	Yes
Unit of fitness maximisation	No	Yes
Cooperation [high] and [low] conflict	Yes	Yes
Bearer of adaptations	Yes	Yes

TABLE 2.1: In the case of the siphonophore *nanomia cara*, many of the putative criteria for identifying biological individuals fail to agree on whether the zooid or the colony is the individual in two ways. Sometimes they do not give a categorical answer based on a single criterion (problem of vagueness) and sometimes different criteria give different answers (problem of disagreement).

categories have blurred edges. The disagreement problem arises from the various attempted definitions picking out different, non-coextensive sets of entities as biological individuals, so they can't all be right. Herron et al. (2013) illustrate both of these problems beautifully with an illuminating series of examples drawn from some of the lesser-studied but major groups of eukaryotes.

An example is informative here. Consider the case of nanomia cara, a curious species of colonial marine invertebrate (animalia/cnidaria/hydrozoa/siphonophorae/nanomia cara). Adult colonies consist of a number of zooids, each of which is homologous to individual and free-living zooids in other hydrozoans. However, like other siphonophores, nanomia cara zooids are not free-living. Colonies consist of many (clonally related) zooids, physically attached to one another and which behave as a functionally integrated whole with division of labour and specialisation of parts. The division of labour includes reproductive division of labour - only a subset of the zooids perform the role of gonads, the others specialising in somatic functions such as pelagic locomotion (swimming), maintaining buoyancy control, catching prey (usually small zooplankton), digesting prey and others (Mackie et al., 1987). If we consider each of the individuation criterion listed above, we find that some of them identify the zooid as the individual, some of them the colony: this illustrates the problem of disagreement. And if we consider any single criterion, there is no guarantee that they will give mutually exclusive answers. This illustrates the problem of vagueness (see Table 2.1).

Any proposed theory of biological individuality, says Clarke (2010), would also need to account for these problems in the previous attempts, to show where they had gone wrong but also why they seem partly right.

2.3 The evolutionary-unit theory of biological individuality

In this section I present what I call 'the evolutionary-unit theory' of biological individuality, which has been gaining momentum in recent years. This theory (or, more accurately, this family of theories) has it that the essential thing about biological individuals is that they are units of evolution (in the sense made precise in Section 1.3). According to this theory, many of the candidate criteria for past attempts to define biological individuality (Section 2.2) can be understood either as factors that either underwrite unit-of-evolution-hood (e.g. spatial contiguity, genetic homogeneity, bottlenecked life cycles) or that are products of it (e.g. histocompatibility, functional division of labour and inter-dependence of parts).

The evolutionary-unit theory of biological individuality is not really a single theory, rather it is a family of positions all sharing a similar evolutionary and hierarchical perspective on biological individuality (Buss, 1987; Maynard Smith, 1988; Godfrey-Smith, 2009; Queller and Strassmann, 2009; Clarke, 2010; Folse and Roughgarden, 2010; Bourke, 2011). The theory exists in a variety of forms with a variety of emphases, depending on the author. I have given it this label as an umbrella term because, I suggest, it has enough coherence to make naming it worthwhile and to make generalisation about it possible. In many ways, the evolutionary unit theory of biological individuality involves very similar theoretical commitments to the major evolutionary transitions research programme, albeit coming from a different starting point and with a slightly different explanatory goal. The major evolutionary transitions research programme seeks to explain 'all stable forms of biological grouping' (Bourke, 2011), while the evolutionary-unit theory of biological individuality seeks to explain why certain classes of objects are (or are not) biological individuals. All versions of the evolutionary-unit theory of biological individuality place a central emphasis on the role of the evolutionary process in shaping individuality. This process is the central explanatory concept that, on this view, holds together and unifies many of the various ways of understanding individuality that have been tried in the past, explaining why they seemed partly right while still suffering from the problem of vagueness and the problem of disagreement.

An evolutionary understanding of individuality is closely connected with a hierarchical view of evolution. Such hierarchical views of the evolutionary process were prominent prior to the formation of the Modern Synthesis in the early decades of the twentieth century[2]. Darwin, in *The Descent of Man* (1871), famously hinted at the idea that a trait might be supported by between-group selection while punished by within-group selection but he did not pursue the consequences of this to any length. Hierarchical thinking is much clearer in the work of Weismann (1903). Pointing out that the basic logic of Darwinism is agnostic about the identity of the units bearing heritable fitness differences, he said that the "extension of the principle of natural selection to all grades of vital units is the characteristic feature of my theories ... this idea will endure even if everything else in [my] book should prove transient' (quoted in Gould, 2002, p.223.)

[2]Bourke (2011, Chapter 1) provides an interesting overview of the historical development of hierarchical evolutionary theory

Ironically, it was not in fact his hierarchical view of the evolutionary process but his doctrine of the 'continuity of the germ plasm' for which Weismann is primarily remembered today. And it is precisely this doctrine, protests Buss (1987), that is responsible for the poverty of hierarchical thinking in mainstream evolutionary biology today. Weismann's doctrine has it that organismal development from the zygote involves a very early forking, such that the heritable material (what he called the 'germ plasm') is segregated from the rest of the body (the 'soma'), lies low most of the time, and is directly responsible for passing on the organism's heritable characters to the next generation. Thus, anything that happens to the soma during the organism's lifetime (due to mutation, physical adaptation, learning or accident) will not be heritable[3]. This (purported) failure of acquired characters to enter the germ line is often known as the 'Weismann barrier'.

At this point I want to define the concept of a **Weismannian organism**. This is an idealisation that will be used as a point of reference in what follows. A Weismannian organism gets fitness by begetting countable, comparable, roughly equivalent offspring that faithfully inherit their heritable characters from their parent(s). Any characters acquired by the parent during its own lifetime are not heritable. Weismannian organisms differ in fitness by differing in (integer) number of offspring (and not by differing in size.) Buss's (1987) book *The Evolution of Individuality* is a landmark text for the major transitions research programme. It gives a modern account of the fact that multicellular individuality is an evolved trait and one that is contingent on the suppression of selection at the cellular level. Buss opens by critiquing mainstream evolutionary biology for generalising from a few genuinely Weismannian organisms to the whole of life, where this idealisation is not at all appropriate:

> "While Weismann's inheritance theories were ultimately proved fictional, their corollary, that the individual is the sole unit of biological organisation, was nevertheless incorporated as a tacit assumption in the modern synthetic theory of evolution. ... The geneticists and naturalists who authored the Modern Synthesis has no pressing reason to raise embryological concerns themselves, as all worked on organisms in which the Weismannian ideal of the individual was closely approximated. ... Weismann's doctrine would be justified, despite its flawed origins, if terminal determination of the germ line always occurred in earliest ontogeny. However, taxa differ in their mode of development. In some taxa, this Weismannian assumption is closely approximated; in others it is not. Crucially, the phyletic distribution of this trait illustrates that early terminal differentiation is a character limited exclusively to some higher metazoan taxa [, i.e. a tiny fragment of the tree of life]. When multi-cellular, cellular-differentiated life first arose, Weismann's doctrine was violated. At this point - and in many taxa even today - it is inappropriate to assume that the individual is the sole unit of selection. Individuality is a derived character." (Buss, 1987, *The Evolution of Individuality*, p.3)

[3]For an interesting account of the historical evolution of diagramatic representations of Weismann's doctrine of the continuity of the germ plasm, see Griesemer and Wimsatt (1989).

As Buss (1987) makes clear, most taxa have life histories involving somatic embryogensis (i.e. much of the soma can potentially beget offspring, not just a special 'segregated germline' part). Most plants and mycelial fungi are examples, among many others (Buss, 1983; Grosberg and Strathmann, 2007). The significance of somatic embryogensis is that it enables within-organismic variation (e.g. mutations) to be inherited by future generations and so enables the operation of within-organismic evolution. I will return several times in this thesis to Buss' point that much of the conceptual foundation of evolutionary theory is implicitly based on the idealisation of the Weismannian organism and that this mistake is responsible for mainstream evolutionary theory's failure to correspond to the actual world in a number of different ways.

Buss' work played a large role in initiating the major evolutionary transitions research programme (Maynard Smith and Szathmáry, 1995; Michod, 1999) and enabled the convergence of thinking on the old problem of biological individuality with more modern concerns about the levels of selection (Lewontin, 1970) discussed in the previous chapter. As Bourke summarises,

> "The concepts of the evolution of individuality and major transitions are themselves underpinned by a key insight ... that the individuality emerging at each major evolutionary transition is a contingent state. Specifically, it is contingent upon the absence or suppression of within-individual conflict .... For, if the level of internal conflict is too great, the higher level of organization either fails to emerge or is unstable and collapses. The challenge has been to understand what kinds of process contribute to the stable evolution of each new level in the hierarchy of major transitions ...." (Bourke, 2011, p.3, Bourke's references ommited)

The move to view biological individuality in the light of evolution is the innovation that led to the evolutionary unit theory of biological individuality.

2.3.1 Godfrey-Smith on Darwinian Individuality

An influential version of what I call the evolutionary-unit theory of biological individuality has been put forward by Godfrey-Smith (2009). His stated goal is not actually to tackle the problem of biological individuality head-on but to address a different question: what properties must a system have if it is to be capable of producing complex adaptations? However, in tackling this question he puts forward what amounts to a version of the evolutionary-unit theory of biological individuality. Godfrey-Smith's approach is to examine what makes a population of things into a 'Darwinian Population' (a population of 'Darwinian Individuals') with the power of adaptation. He places great emphasis on the idea that Darwinian Individuality is a matter of degree and is not an all-or-nothing category. The analysis is first presented on a single hierarchical level, giving a multi-dimensional account of what makes a population a Darwinian Population (2009, Chapter 3). He then moves to a multi-level analysis, discussing 'collective reproducers' (2009, Chapter 5) akin to the particles and collectives discussed in this thesis (Figure 1.1), where he

gives a similarly multi-dimensional account of what makes it the higher or the lower level that has the most Darwinian Individuality. (In subsequent chapters I will use the term 'social niche modifiers' to refer to many of the phenomena Godfrey-Smith casts as determinants of Darwinian Individuality. Recall that there is a Glossary at the back of this thesis, just before the references.)

On a single level, says Godfrey-Smith, a population of reproducing entities can vary in the extent to which it is Darwinian along a number of (partly but not entirely orthogonal) dimensions, including:

- H - fidelity of heredity
- V - abundance of variation
- S - dependence of fitness variance on intrinsic rather than extrinsic properties
- and several others

For example, the cells in a human body have high fidelity of heredity (H) because DNA replication at mitosis is faithful, but very low abundance of variation (V) because they are all clonaly related, somatic mutations notwithstanding. And variance in their long-term fitness does not depend on their intrinsic properties, (such as genome) but on their physical location in the body. Cells that happen to be in the gonads can potentially have non-zero direct fitness, while cells in all other organs are evolutionary dead ends that are certain to have no long-term direct offspring. Therefore, the cells in a human body are not Darwinian Individuals.

The very continuousness of these properties (and the fact that, although sometimes connected, they can vary independently of one another) means we are talking about *individuality as a matter of degree*. This is very important in itself - previously, failure to give a categorical answer about individuality was taken to be a problem (see 'the problem of vagueness' in Section 2.2 above).

Of particular interest to my project in this thesis, Godfrey-Smith (2009, Chapter 5) discusses a category of Darwinian Individual he calls 'collective reproducers'. These are entities that reproduce and that are made of parts that also reproduce, along the lines of the collectives and their constituent particles (illustrated in Figure 1.1) that feature prominently in this thesis. Multicellular organisms, colonial organisms like corals, and eusocial societies like leaf-cutter ants or honey bees are good examples.

Part of Godfrey-Smith's agenda is to attack the idea that Darwinian Individuality is a clear-cut category. Just as in his single-level analysis, he argues that collective reproducers are Darwinian Individuals in their own right to a greater or lesser extent, depending on their possession of certain properties. This is directly analogous to my interest in what makes collectives into units of evolution (Section 1.4). For Godfrey-Smith, collectives are Darwinian Individuals and not mere groupings of particles when they score highly on the following dimensions:

- B - life-history bottleneck

- G - germline segregation
- I - functional integrity of parent and offspring as distinct units

For example, honey bee (Apis mellifera) colonies exhibit a life-history bottleneck (B), because a single singly-mated queen founds the colony and lays almost all the eggs. Such a population structure also embodies a germline segregation (G) - there is reproductive division of labour (among females, at least) in that the workers' egg-laying potential is hormonally suppressed by the queen and that worker-laid eggs are likely to be destroyed by other workers. There is also functional integration of the colony (I) due to non-reproductive division of labour and specialisation of roles in different worker castes. Colonies have a clear life-cycle with a founding event and a reproduction event - they do not come into being through simple growth of pre-existing colonies. All this means honey bee colonies score highly as Darwinian Individuals on Godfrey-Smith's multi-dimensional analysis. Communal sweat bees (Agapostemon virescens), on the other hand, do not. They live communally but their communities lack a life-history bottleneck and germline segregation (all females can reproduce), and they do not exhibit functional division of labour or specialisation of roles.

Godfrey-Smith's (2009) multi-dimensional approach to individuality is superficially very different to my character-fitness covariance approach, yet quite consistent with it. This is because Godfrey-Smith's dimensions along which populations of entities vary in their Darwinian Individuality are *realizers* of character-fitness covariance. To see why, consider the following.

Start with the single-level account of what makes a population into a Darwinian Population. Recall from Section 1.3 that a population will show evolutionary change when there is non-zero covariance between its members' fitnesses, w, and the mean character of their offspring, z'. That is, $|cov(w, z)| > 0$ and $|cov(z, z')| > 0$ so that $|cov(w, z')| > 0$. Godfrey-Smith's fidelity of heredity dimension - the extent to which like begets like - can be seen as a way of thinking about the $|cov(z, z')| > 0$ part of the character-fitness covariance formulation. Similarly, his abundance-of-variation dimension can be understood as impinging on the $|cov(w, z)| > 0$ part. This is because there must be variance for there to be covariance, so a necessary condition for $|cov(w, z)| > 0$ is that $|var(w)| \neq 0$. Again, Godfrey-Smith's dimension concerned with the dependence of fitness variance on intrinsic rather than extrinsic properties can be understood as saying something about the meaning of the terms in the $|cov(w, z)| > 0$ condition. It says that z must be a property that inheres in the entity and not an accidental property (genetically determined properties would 'inhere in the individual' in this sense).

Now consider Godfrey-Smith's two-level account of Darwinian Populations. This, also, can be understood as a discussion of various properties that act as realizers of character-fitness covariance at collective level, or suppressors of it at lower levels. (I'm using uppercase letters for collective fitness W and character Z and lowercase letters for particle fitness w and character z in this paragraph.) Life-history bottlenecks (B) deny abundance-of-variation to the particle population, leaving most variation between collectives rather than within them. That is, they can

be understood as ensuring that $|var(W)| \neq 0$ and that $|var(w)| = 0$. Germline segregations (G) ensure that fitness differences between particles are due to differences in extrinsic properties (such as which side of the reproductive division of labour they find themselves) rather than intrinsic differences (such as genetic properties). This can be understood as a way of saying that $|cov(w, z)| = 0$ at particle level when a germline segregation is in force. Both of these are about strengthening Darwinian Individuality at collective level by denying it to the particle level below. Godfrey-Smith's third criterion for assessing the Darwinian Individuality of collective reproducers - functional integration (I) is the least amenable to a character-fitness covariance reading. Perhaps this is related to some explanatory cart-before-horse worries I will return to later; there is a sense in which division of labour - both reproductive (G) and functional (I) - is a consequence rather than a cause of Darwinian Individuality. By that, I mean that these divisions of labour are themselves evolved characters, rather than pre-requisites for evolution to occur[4]. I postpone discussion of this point until the final chapter.

So far, my review of Godfrey-Smith's theory of Darwinian Individuality has been synchronic - I have discussed his analysis of what makes a population be a Darwinian Population but not how things came to be like that. Godfrey-Smith does not present a detailed diachronic theory for evolutionary transitions in individuality but he introduces some useful terminology to describe them, notably the concept of 'de-Darwinization' (Godfrey-Smith, 2009, Section 5.3). This is a diachronic concept (sensu Okasha (2005b)). Godfrey-Smith suggests that, for Darwinian Populations to be nested, one level must 'de-Darwinize' the other. In the following section on Clarke's views, this idea is made much more clear, particularly the reason why it must be a hard tradeoff. De-Darwinization is therefore analogous to the process I take as my explanandum in this thesis (Section 1.4).

While I find much of Godfrey-Smith's analysis consistent with my own, there are also points of disagreement. Part of what Godfrey-Smith is doing with his analysis of collective reproducers is to find some principled way to distinguish between reproduction (production of a new individual by a pre-existing one) and growth (enlargement of a pre-existing individual), because he assumes that only bona fide reproduction yields fitness and so Darwinian Individuality. I differ from Godfrey-Smith on this point. In chapter 6 I will argue that there is a sense of fitness - the sense important for evolutionary change in the actual world rather than in abstract models - for which differential reproduction is not a necessary ingredient (differential growth and differential persistence will also do the job). Therefore I do not share Godfrey-Smith's motivation for distinguishing reproduction from growth.

[4]One of the criteria that has historically been used in attempts to define biological individuality has been to say that biological individuals are the bearers of adaptations (Section 2.2). They are, but this is a consequence rather than a cause of their individuality because, as Williams (1992) stresses, adaptation is evidence of past evolution and not a pre-condition of it.

2.3.2 Clarke on biological individuality and character-fitness covariance

As with Godfrey-Smith, Clarke's work on biological individuality has had a significant influence on this thesis (Clarke, 2010, 2012, 2013, 2014). An important strand in Clarke's position is that the previous attempts to define biological individuality (discussed in Section 2.2 above) failed because they did not recognise that individuality is a multiply realizable functional role, where that role is 'being a unit of evolution'. Historic attempts have mistaken the contingent for the essential - they have all tried to define biological individuality in terms of factors that are realizers of level-of-selection-hood rather than in terms of level-of-selection-hood itself. Although Clarke does not make this criticism of him, it is clear that Godfrey Smith also fails to recognise (or stress sufficiently) that his list of determinants of Darwinian Individuality are all contingent ways to arrive at level-of-selection-hood - and that this is what is important. So past attempts all mistake ways of getting to what's really important from the really important thing itself.

Where Godfrey-Smith's mixed bag of properties conveying individuality is self-consciously pluralist and messy, Clarke's approach brings some conceptual unification. By separating a multiply realisable functional role from the ways in which it is realized, she is able to account for the huge amount of disagreement among the many previous attempts at defining biological individuality, while still allowing that each of them did seem to be onto something important and not totally wrong. Furthermore, given that the strength of the evolutionary process is a matter of degree, the problem of vagueness discussed in Section 2.2 can also be explained away.

2.3.3 Modeling multi-level biological individuality with the multi-level Price Equation

While Godfrey-Smith argued verbally for a multi-level understanding of individuality, Clarke (2010, Chapter 4) makes this idea much more precise with what she calls the Price Approach to biological individuality. Clarke uses a multi-level version of the Price Equation (Price, 1970; Okasha, 2006; Gardner, 2008) to quantify the extent to which character-fitness covariance is present at some level in a multi-level hierarchy, relative to the extent to which it is acting at other levels (e.g., to borrow Sober's (1984) phrase, relative to 'the group above and the gene below'). In a two-level hierarchy like the one used in the definition of the thing to be explained in this thesis (Section 1.4), a two-level Price Equation can be used to measure the relative strengths of particle-level and collective-level character-fitness covariance and then to compare them.

Metaphysical doubts about causation and correlation aside, the two-level Price Equation can be understood as partitioning change into that due to selection on particle characters and that due to selection on collective characters. Following Okasha (2006, Section 2.3.1):

$$\underbrace{\bar{w}\Delta\bar{z}}_{\text{total change}} = \underbrace{E(cov_k(w,z))}_{\text{change due to particle level selection}} + \underbrace{cov(W,Z)}_{\text{change due to collective level selection}} \tag{2.3.1}$$

This assumes all change is due to selection (i.e. that particles breed true). The particle-level selection term is the mean of all the collective-wise covariances of particle character with particle fitness. There is one $cov(w, z)$ value per collective (collectives indexed by k). This then yields a single value of $E(cov_k(w, z))$ for particle-level selection in the population as a whole. The collective-level selection term is $cov(W, Z)$, where collective fitness W is mean particle fitness and collective character Z is mean particle character. The covariance function takes one pair of (W, Z) values per collective and yields a single value for collective-level selection in the population. Both measurements are made over the same time scale (perhaps a single generation).

Clarke's Price Approach to biological individuality in a two-level setting is to compare the total change with that due to the first and second terms here. The relative strength of collective-level selection is then:

$$\frac{|cov(W, Z)|}{|cov(W, Z)| + |E(cov_k(w, z))|} \tag{2.3.2}$$

Absolute values for the covariances are used, because we want to compare their relative magnitudes without regard to their direction. Relative collective-level selection plus relative particle-level selection adds up to 1. Clarke (2010) uses the Price Approach to emphasise a philosophical point about the potentially multi-level nature of individuality at a given time, rejecting the assumption that it is an all-or-nothing category that applies to only one level at a time.

The important distinction between units of selection and units of evolution was discussed in Section 1.3 above. I introduce the following modification to Clarke's version of the Price Approach. To capture selection and heredity, $E(cov_k(w, z'))$ and $cov(W, Z')$ are used (where z' and Z' are particle and collective offspring characters respectively.) This allows the Price Approach to be used to measure the extent to which entities at a level are units of evolution (and not just units of selection, as in Clarke's version). The relative strength of collective-level evolution is then:

$$\frac{|cov(W, Z')|}{|cov(W, Z')| + |E(cov_k(w, z'))|} \tag{2.3.3}$$

I have set up the thing to be explained about ETIs in terms of how character-fitness covariance gets shifted from particle to collective level by an evolutionary process acting initially at particle level (Section 1.4) in a population having a two-level hierarchical structure (Figure 1.1). The Price Approach allows a quantitative analysis of the level at which the evolutionary process is acting in any such model scenario. I put the Price Approach into practice later in this thesis. In Chapter 4 I apply it to a simulation model, to measure the strength of the evolutionary process at two levels diachronically - observing how it changes during a simulation run and demonstrating an evolutionary transition in individuality.

2.4 MLS1, MLS2, fitness$_1$ and fitness$_2$

The evolutionary-unit theory of biological individuality is explicitly hierarchical, acknowledging that some population of evolutionary units are made of parts that are also evolutionary units. This invites a multi-level formulation of evolutionary theory, usually known as 'multi-level selection' theory.

Throughout this thesis, reference will be made to two different understandings of the term 'multi-level selection', which I will call MLS1 and MLS2 (Damuth and Heisler, 1988; Okasha, 2005b). Both have been employed at different times and in different places in the literature on group selection, the evolution of cooperation, evolutionary transitions in individuality and associated problems. The distinction between them will be important in what follows. For the reader unfamiliar with this distinction, this section presents a quick primer on the two abstract processes MLS1 and MLS2 and the two related properties fitness$_1$ and fitness$_2$.

Both MLS1 and MLS2 are abstractions that involve a group-structured population with particles nested inside collectives, as in Figure 1.1. In MLS1, particles are the focal units, while in MLS2 both particles and collectives are tracked.

In MLS1 models, we are interested in the evolution of particle properties in the global particle population; the collective structure is simply part of the particles' environment. Collectives form, particles go about the business of their reproduction within them and the collectives then dissolve, returning the particles to an unstructured pool from which they join new collectives to form the next generation. Particles are units of evolution in these models (in the sense elaborated in Section 1.3), collectives are not. In MLS1 models, collectives do not participate in the parent-offspring relation and are not the sorts of things of which heritability can be predicated. In these models, collectives do not directly beget other collectives and so there is no sense of collective-level fitness, understood as the number of offspring collectives begot by a focal collective (call this collective fitness$_2$). However, we can still track how many particles each collective contributes to the global particle population (call this collective fitness$_1$) and so there is still a sense in which some collectives can fare better than others.

In MLS2 models, both particles are collectives are tracked. Particles go about the business of their reproduction within their respective collectives as with MLS1. However, MLS2 models explicitly represent collectives as evolutionary units in addition to particles. Collectives participate in the parent-offspring relation, directly begetting other collectives. Call collective-level fitness in this sense collective fitness$_2$. In MLS2 models, collectives possess heredity to the extent that like collectives beget like.

It is natural to assume that fitness$_1$ is a measure from MLS1 models and fitness$_2$ a measure from MLS2 models but this is not quite right. There is no reasonable role for a fitness$_2$ measure in MLS1 models. However, the notion of fitness$_1$ - a measure of how many particles a collective contributes to the global particle population - still makes sense in MLS2 models, in addition

to fitness$_2$. (I will appeal to just such a sense of collective-level fitness in an MLS2 context in Chapter 6, in discussing Van Valen's views on fitness and how they can help us with some problems I diagnose in contemporary theory on evolutionary transitions in individuality in Chapter 5.)

Both abstractions, MLS1 and MLS2, have a long history in the literature (often without those names attached). MLS1 models are invoked in many well-known models from the group selection literature, including Williams and Williams's (1957) model for the evolution of altruism in sib groups, Maynard Smith's (1964) 'haystack model' and D.S. Wilson's (Wilson, D.S., 1975) 'trait group model'. For a review, see Sober and Wilson (1998). An MLS1 model was employed by Powers in early work on social niche construction (Powers, 2010; Powers et al., 2011). De Monte and Rainey (2014) suggest an MLS1-like model can be used to think about the evolutionary dynamics of Dictyostelium discoideum (cellular slime mould), with the cells in the role of particles (showing a clear parent-offspring relation) and the slugs in the role of collectives (without any clear parent-offspring relation).

MLS2 models have often been used in the context of speciation, species selection and paleontology (Van Valen and Sloan, 1966; Vrba, 1984; Gould, 2002). The idea here is that, in addition to the evolutionary process acting upon populations of conspecifics, species themselves also beget offspring (new species) that resemble their parents - and do so at differing rates, resulting in the extinction of some and the increase of others. The evolutionary process therefore operates on both levels simultaneously.

A prominent theory (Fitness Export Theory) belonging to the major evolutionary transitions research programme, due to Michod (1999) and discussed further in Chapter 5, draws on both types of model, suggesting that MLS1 is involved early in a major transition while MLS2 is implicated in the latter stages. More will be said of this in due course. The goal of this section is to make clear the meaning of the terms MLS1, MLS2, fitness$_1$ and fitness$_2$.

Chapter 3

Social Niche Construction

3.1 Overview of chapter

I explain what social niche construction is and how it can help explain evolutionary transitions in individuality.

I characterise the standard explanatory model in social evolution theory as adopting an externalist explanatory stance (sensu Oyama 1992, Godfrey-Smith 1996). Factors like relatedness, reciprocity or co-dispersal are cast in the role of 'external factors' and social behaviour (often cooperation of some sort) is cast in the role of 'adaptive response' to that external environment. Lewontin (e.g. 1985) famously challenged externalism about biological adaptation in general, arguing that organisms and their environments co-define and co-construct one another. I bring those ideas to bear on the debate about social evolution theory and the evolution of cooperation. In this spirit, I adopt and develop Powers' (2010; 2011) theory of 'social niche construction' (SNC). The central idea is that organisms partly construct the *social* niches in which they live; they can change the game being played, sometimes making conditions more favourable for cooperation (and thereby increasing in absolute fitness). That allows me to take an interactionist explanatory stance in which social behaviour and the above factors modifying the social niche (relatedness, or reciprocity, or others) can be modelled as reciprocal responses to one another.

While I start by talking in very general terms about the social niche as the 'selective context in which social behaviour occurs' and about organisms being 'both the subject and object of their own social evolution', I go on to give a much more precise account of these things for the sub-class of cases involving social evolution among like-kinds (e.g. coloniality, multicellularity and the fraternal evolutionary transitions)[1]. I borrow some formalism from evolutionary game theory, using it as a precise way to describe a social niche - the strength and direction of selection on social behaviour. A minimal two-locus, two-allele model of social niche construction is

[1]This class was selected because it is more amenable to analysis (by me) than social evolution among unlike-kinds (e.g. inter-specific mutualisms and egalitarian evolutionary transitions).

sketched out. One locus is expressed as a Social Trait (ST), tentatively 'cooperate' / 'defect' in a collective action problem. The other is expressed as a Social Niche Modifiying trait (SNM). (There is a glossary of terms in Appendix A, just before the references.)

Examples of social niche modifiying traits include old favourites such as life-history bottlenecks, early-segregating slowly-dividing germlines and worker-policing in social insects. However, rather than special cases or rare curiosities, I follow Bourke (2011) in interpreting a very wide range of structural features of the biological world as playing the functional role of social niche modifiers, ubiquitous wherever complex adaptation is in evidence (e.g. the self/nonself discrimination mechanisms in filamentous fungi, the apical meristem topology of vascular plants and the obligate co-dispersal of endophytic fungi with their symbiotic sedges). Table 3.1 provides many examples. Social niche modifiers quantifiably change the effective game being played between particles, moving particle and collective fitness interests into greater (or lesser) alignment.

What is new here is not the idea *that* collective life-forms have features that ameliorate or avoid internal conflict (Michod and Herron, 2006; Queller and Strassmann, 2009) but to view the *evolution* of such phenomena as admitting of a general theoretical treatment.

3.2 Internalist, externalist and interactionist explanatory stances

Explanations for the properties of organisms can usefully be understood as belonging to one of a number of broad categories, including internalist, externalist and interactionist (Oyama, 1992; Godfrey-Smith, 1996). This applies to both their ontogeny and their phylogeny.

Internalist explanations appeal to circumstances within the organism whose observable properties are to be explained. For example, an explanation for the properties of a particular organism might be given in terms of the organism's genetic properties and an unfolding or revealing of those pre-existing properties by the developmental process (e.g. Jacob, 1970). Such an explanation is internalist to the extent that it emphasises internal causes or influences on development at the expense of environmental ones. It is also possible to give an internalist explanation for the evolutionary trajectory of a population of organisms. An explanation for the properties of a population that cites developmental constraints limiting the range of possible phenotypic variation would be internalist in this sense (e.g. Gould and Lewontin, 1979; Maynard Smith et al., 1985). Kitcher (1985, pp. 214-226) provides several detailed examples of cases involving diploidy and overdominance where selection imposed on a population by its environment cannot select the fittest combinations of existing alleles, due to internal constraints arising from the nature of diploidy itself. This sort of internalism is not absolute - it amounts to an acknowledgement that there are limits on the explanatory power of external selection. Other internalist explanations place internal factors centrally and not just as modifiers or constrainers of external factors. When Kimura (1968; 1983) argued that the principle agent of evolutionary change is not natural selection but random drift among selectively-neutral alternatives, he was advocating an internalist position. Lamarck was an internalist about adaptation, stressing the role of internal striving

in his theory of the shaping of the characteristics of organisms. For Lamarck, change in a population results from the aggregation of internally-driven changes in the individuals comprising it (Lewontin, 1985, pp.85-89).

Opposed to internalism, the externalist explanatory stance casts external factors in the primary causal role. Adaptationism is a classic example of an externalist explanatory position in biology. On this view, the properties of populations of organisms are determined by factors largely or entirely external to the individuals in them (Spencer, 1864; Lack, 1947; Simon, 1981; Williams, 1966; Brandon, 1990; Godfrey-Smith, 1996). The evolutionary process here is one of moulding to or shaping of organisms by their environment. In many cases, externalist explanations do allow room for internally-located constraints (perhaps developmental or genetic) but they lend much greater weight to consideration of external factors. Clearly, the environment in which organisms live is in some ways influenced by their presence and action in it and moderate forms of externalism do not deny this, although they may not give it much attention. A stronger form of explanatory externalism ('asymmetric externalism' in Godfrey-Smith's 1996 terminology, p.132) holds that external factors shape the properties of organisms and that no feedback occurs in the other direction. The claim here is that, while environments may change for their own reasons, organisms do not influence their environments in a way that changes the manner in which those environments impact on those organisms' evolution. This asymmetric externalist stance is captured succinctly by Williams (1992, p.484) when he says "Adaptation is always asymmetrical; organisms adapt to their environment, never vice versa."

Roughly, the internalist stance has been dominant in developmental biology while the externalist stance has been dominant in evolutionary biology. Lewontin (1985) attributes this historical contingency to the influences of Mendel and Darwin respectively.

Godfrey-Smith (1996, p.132) offers several different ways an asymmetric externalist explanation can go wrong. It might be the case that organismic properties are best explained in an internalist way (and the same could apply to the properties of the environment). In this scenario, there may be no explanatory link between the properties of the environment and the properties of the organism. Alternatively, it might be the case that it is the properties of organisms that explain the properties of the environment, a reversal of the explanatory arrow in the asymmetric externalist story. Lewontin (1985) calls this a 'constructivist' pattern of explanation. For example, the evolution of photosynthesis in the predecessors of the cyanobacteria had the effect of dramatically affecting the composition of the earth's atmosphere (Lenton and Watson, 2011). An explanation for how the atmosphere came to be oxygenated (a thermodynamically unstable state standing in need of explanation) citing the effects of that particular form of autotrophy would be constructivist in this sense.

A third possibility is that the properties of organisms can be explained by properties of their environments *and* that the properties of environments can be explained by the properties of the organisms in them (at least partly, in both directions). This is the type of explanation I will develop in this thesis. Oyama (1985, 1992, 2000) calls this type of explanation 'interactionist', not

only to distinguish it from internalism and externalism but also to de-dichotomize the debate. She first discussed interactionism in the course of her work on developmental systems theory, to refer to "a way of thinking about development that does not rely on a distinction between privileged, essential causes [genetic ones] and merely supporting or interfering causes [environmental ones]" (2001). However, interactionism is also an explanatory stance one can adopt about evolutionary theory and many have done so - notably Lewontin (1985) and Odling-Smee et al. (2003). In this thesis, I apply an interactionist stance to explanations in *social* evolution theory - the so-called 'evolution of cooperation' and the related research programme on major evolutionary transitions.

Interactionism can be a slippery term. While Oyama's interactionism had its origins in the context of nature/nurture debates about development, the term has subsequently come to be used in many overlapping but different ways in many other contexts where the debate between internalist and externalist positions is found (including evolutionary theory). Consequently it carries some danger of misinterpretation, so I will explicitly state that 'constructivist interactionism' is the version I intend. The key insight is that evolution is a constructive process: "evolution is not a matter of organisms or populations being molded by their environments, but of organism-environment systems changing over time" (Oyama et al., 2001, p.2). Although he used a different set of terminology derived from Marxist dialectics, Lewontin is well known for arguing an essentially constructivist interactionist position (1982; 1985; 1991; 2000). See also Godfrey-Smith (2001) for further discussion of Lewontin's views.

> *"It is impossible to avoid the conclusion that organisms construct every aspect of their environments themselves. They are not the passive objects of external forces, but the creators and modulators of those forces. The metaphor of adaptation must therefore by replaced by one of construction, a metaphor that has implications for the form of evolutionary theory." (Lewontin, 1985)*

Lewontin illustrates his point quite neatly with some simple differential equations that have subsequently been adopted by several other authors (e.g. Godfrey-Smith, 1996; Odling-Smee et al., 2003). He diagnoses the adaptationist (i.e. externalist) pattern for an evolutionary explanation as to why an organism (or population of them) has the properties it does (O) in a particular environment (E), as follows:

$$\frac{dO}{dt} = f(O, E)$$

That is, the properties of the organism (or population thereof) at a time depends on the state of the population at a previous time and the state of the environment. E explains O. The environment is taken to go its own way, perhaps changing but not as a function of the organism:

$$\frac{dE}{dt} = g(E)$$

Lewontin argues that this model is misconceived because it fails to give due consideration to the effects organisms have on their own selective environments. His constructivist-interactionist

position is then illustrated by coupling the equations, so that organism and environment each change in response to the other:

$$\frac{dO}{dt} = f(O, E)$$

$$\frac{dE}{dt} = g(O, E)$$

"The coupled differential equations that describe their coevolution are not easy to solve, but they represent the minimum structure of a correct theory of the evolution of such systems. It is not only that they are difficult to solve, but that they pose a conceptual complication, for there is no longer a neat separation between cause (the environment) and effect (the organism). There is, rather, a continuous process in which an organism evolves to solve an instantaneous problem that was set by the organism itself, and in evolving changes the problem slightly.
... the organism influences its own evolution, by being both the object of natural selection and the creator of the conditions of that selection." (Lewontin, 1985, pp.105-106). So we arrive at a model of explanation something like 'E explains O explains E'. See also Sterelny and Griffiths (1999, 11.4), Barberousse et al. (2009) and Godfrey-Smith (2014, pp.54-59) for constructive-interactionist discussion of the relation between organism and environment or niche.

3.3 Social evolution theory usually employs an externalist explanatory stance

Social evolution theory typically employs an externalist explanatory stance. This sort of explanation starts by noting the existence of some form of biological cooperation or altruism. This behaviour is taken to stand in need of explanation because, on first examination, it is inconsistent with the predictions of evolutionary theory - Darwin's 'one special difficulty'. Cooperation is paradoxical for evolutionary theory, until we notice the presence of some factor that explains why the observed behaviour is adaptive after all. In many well-studied cases, the factor used to explain the social behaviour is genetic relatedness between actor and recipient (Hamilton, 1964). In others, it is explained by the presence of iterated interaction between re-identifiable individuals (Trivers, 1971; Axelrod and Hamilton, 1981). In still further cases, among symbiotic mutualists, it is explained by vertical transmission (co-dispersal) (Ewald, 1987) or partner discrimination (Noë, 2001; Kiers et al., 2011; Archetti et al., 2011). The explanatory pattern is linear and proceeds in only one direction: there is some factor external to the organism that explains some social trait of the organism as an evolutionary response to that factor. In the absence of the external factor, we would not expect cooperation to be evolutionarily stable. In the presence of the factor, we understand cooperation to be an adaptive response to it. Different classes of cooperation are explained by different classes of external factors. For example, within-colony kinship might be used to explain reproductive altruism in eusocial hymenoptera (Bourke and Franks, 1995), while life-history considerations pertaining to vertical transmission are used to explain ongoing stability of mutual cooperation in the symbiosis between the

aphid Acyrthosiphon pisum and its bacterial endosymbiont Buchnera aphidicola (Wernegreen and Moran, 2001). Different cooperative or altruistic behaviours have different explanations and those explanations each invoke some external factor which is itself unexplained.

The externalist explanatory stance is not wrong but it can be problematic in that it guides our thinking in directions that are not always fruitful. The use of an externalist explanatory stance has proven particularly problematic for attempts to explain evolutionary transitions in individuality using the tools of social evolution theory. I will return to this topic in the final chapter of the thesis (Section 7.2).

3.4 A general Darwinian explanation for the origin of factors enabling cooperation

The factors invoked to explain cooperation[2] are rich and varied. One might ask whether a general Darwinian explanation for the *origin* of those factors can be given. This line of inquiry could go one of a few ways. We might find that, in each case, the factors to be explained arose in a manner for which no Darwinian explanation is (or is yet) available. Alternatively, we might find that, while each case admits of Darwinian explanation, all the cases are very different. Perhaps there are no general principles to be discovered about the evolution of these things and the best we can hope for is a 'patchwork' of explanations (Dupré, 1995; Cartwright, 1999). After all, explanations involving relatedness, reciprocity and group selection all look rather different. Or perhaps, as I contend, a general explanation is available - one that offers conceptual unification across the apparently disparate cases. That is the motivation for my use of social niche construction theory.

3.5 SNC employs an interactionist explanatory stance, reciprocally explaining both cooperation and the factors enabling it

> *"...the organism influences its own evolution, by being both the object of natural selection and the creator of the conditions of that selection."*
> *(Lewontin, 1985, p.106)*

Social evolution theory tries to explain the evolution of social behaviours (Hamilton, 1964; Wilson, E.O., 1975; Bourke, 2011); niche construction is any process in which organisms modify their own environment in such a way as to influence the conditions of their own evolution (Odling-Smee et al., 2003; Laland and Sterelny, 2006). The term 'social niche construction' denotes the application of niche construction theory to social evolution (Powers, 2010; Powers et al., 2011).

[2] In what follows, I will call these factors 'social niche modifiers'.

In the dialectical spirit of Lewontin (1985) and the Niche Construction theorists (Odling-Smee et al., 2003), social niche construction theory employs an explicitly interactionist explanatory stance (Oyama et al., 2000, 2001; Barberousse et al., 2009). The central idea is that biological entities are both the subject and object of their own social evolution. The advantage of this sort of thinking is that it permits the explanation of changes in terms of their effects without running into any metaphysical problems about the order of cause and effect, it avoids explanatory postponement and it dissolves the apparent paradox facing attempts to explain how natural selection acting at some level in the biological hierarchy could shift that very process to a higher level from below. I suggest the interactionist explanatory stance corresponds more closely to the causal structure of the biological world in this context.

Some terminology:

- **particle, collective** - lower and higher level entities in a two-level part-whole hierarchy (following Okasha, 2006).
- **social trait (ST)** - a trait that affects the fitness of individuals other than the actor (e.g. Kropotkin, 1902; Bourke and Franks, 1995; Crespi, 2001; Calcott, 2008), sometimes having values appropriately labelled 'cooperate' or 'defect' .
- **social niche** - the selective context in which social behaviour occurs, affecting the strength and direction of selection on it. In game theoretic terms, the social niche is the effective game being played, once all relevant factors have been taken into account. (Relevant factors include the underlying game, any social niche modifiers (SNMs) that may be present and the frequencies of various social traits in the population).
- **social niche modifier (SNM)** - a trait that alters the effective game being played by its bearers, causing it to differ from the counterfactual game they would have been playing if the social niche modifier had not acted. Examples include factors such as population structure, relatedness, punishment, policing and side-payments. See Table 3.1 on page 38 for many examples - they make it a lot clearer.
- **social niche construction** - a circular process in which organisms modify their own social niche in such a way as to influence the conditions of their own social evolution.

Social niche and social niche modifier are inter-defined. The social niche modifier trait is called a 'modifier' because it changes the social niche that *would have* obtained had it not been in operation. It is by the action of social niche modifier traits that individuals can (partly) construct their own social niche and so influence the circumstances of their own social evolution. Policing, punishment, side-payments and relatedness among interaction partners are examples of social niche modifiers that can arise as a result of individuals' social niche modifying traits. (Michod (1999, p.137) calls them 'conflict mediators'). It is important for their evolution that social niche

modifiers alter the effective game being played for their bearers' interactions, not for the whole population[3]. That is, social niche modifiers assort[4].

[The two-trait model used here is inspired by the one used in the niche construction literature (Odling-Smee et al., 2003) and earlier work on social niche construction (Powers, 2010). Niche construction theory uses the terms 'niche constructing trait' and 'recipient trait'. I rename the recipient trait 'social trait' because I am dealing with social evolution in particular. Why do I label the analogue of the niche constructing trait 'social niche modifying trait', rather than 'social niche *constructing* trait'? Following Lewontin (1985), niche construction theory explicitly challenges the idea that niches exist independently of their occupants. Rather, niches and their occupants are seen as co-defining and co-constructing one another. However, when it comes to social evolution there is a sense in which there *is* a pre-existing social niche that obtains in the absence of any social niche modification - usually a zero sum game or at best an unmitigated Tragedy of the Commons (both amount to selfish behaviour being selected-for). My choice of terminology was motivated by the intuition that, in a finite world, a social niche red in tooth and claw needs no construction. Conversely, niches that support cooperation of some sort stand in need of explanation and, I contend, that explanation involves social niche *modifiers* that modify the social niche that would have existed had they not acted. That is why my terminology differs from that used in niche construction theory more generally.]

Assortment can be an important social niche modifier. Consider a situation where a population faces a public goods game instantiating a Prisoners' Dilemma and is freely-mixed. Here the social niche is a Prisoners' Dilemma, where defection is the evolutionarily stable strategy (ESS). Now consider a contrasting situation where a population faces the very same public goods game but in the presence of a population structure, such that individuals interact only with clones of themselves. The effective game being played in this modified social niche is a Harmony Game, where cooperation is ESS. I return to discuss assortment as a social niche modifier at greater length in Section 3.8.3 below.

In another example, consider punishment as a social niche modifier (Boyd et al., 2010). Suppose again that a population faces a public goods game instantiating a Prisoners' Dilemma. In the absence of any social niche modifier, the social niche is a Prisoners' Dilemma and defection is the evolutionarily stable strategy. In the presence of punishment as a social niche modifier, any

[3]See Jackson (2015); Jackson and Watson (2016) for a general mathematical treatment of the effect of SNM-assortment on social niche construction. In short, they show that social niche construction can modify any base game in the direction of greater cooperation when there is full SNM assortment and can modify only a subset of base games in the direction of greater cooperation when there is no SNM assortment. In the biological world, assortment of SNM often arises quite easily. For example, if we allow that particles within a collective are more likely to be spatially proximate to their closer relatives (e.g. due to cell division with limited dispersal) then it is plausible that particles included in a propagule would be similar to one another in that regard.

[4]One might worry that the assumption that social niche modifiers assort simply postpones the problem of explaining how cooperative behaviours assort. However, the assumption of SNM assortment is benign. The special difficulty explaining social trait assortment is that, in the absence of any social niche modifiers, a social dilemma obtains. There is short-term selective disadvantage on any player cooperating, so positive assortment of cooperative behaviour is not evolutionarily stable. However, there is nothing intrinsically maladaptive about SNM assortment. It does not cause any short-term change in payoff one way or another (and it can occur quite innocently and for no special reason e.g. budding off with stickiness yields symmetrical relatedness between neighbours.)

temptation to defect is reduced by probable punishment, wiping out any gain that might be made from unilateral defection. In this social niche, cooperation is ESS even though defection would have been ESS if the social niche modifier of punishment had not been in force.

Recapping, the social niche is the *effective game* being played, once relevant factors have been taken into account. There is an important distinction between the social niche actually encountered and the social niche that would have been encountered in the absence of those relevant factors. A large number of putative examples of social niche modifiers are provided in Table 3.1 on page 38 below (and the concept is much clearer in the light of some biological examples).

Social niche construction is a circular process in which organisms modify their own social niche in such a way as to influence the conditions of their own social evolution. If a population varies in its social niche modifying trait then it is possible that not all individuals in the population experience the same social niche. If some focal set of individuals (minimally two) locally modify their social niche in a pro-social manner, then this can yield a change in the level of cooperation among those focal individuals (raising it above the level of cooperation among the wider population). The benefits of this increased cooperation increase the fitness of the bearers of the social niche modifying trait. (Note that this is not because there is direct selection on the social niche modifier but because it is correlated with the social trait - and the social trait, in the locally modified social niche, confers a fitness advantage on the bearer.) There is a circularity here that warrants emphasis:

- a pro-social allele at the social niche modifier (SNM) locus enables higher levels of cooperation at the social trait (ST) among its bearers and selection responds to this
- a higher level of cooperation (ST) leads to higher fitness for the bearers of the pro-social allele at the (linked) social niche modifier (SNM) locus
- repeat

In this way, runaway selection on a linked pair consisting of an initially-rare mutant pro-social allele at the SNM locus and an initially-rare mutant cooperative allele at the ST locus can potentially invade a population bearing a wild-type SNM allele that does not enable cooperation.

Social niche construction theory predicts that whenever we find cooperative behaviour in the biological world, we expect to find co-evolved mechanisms supporting it. Without the mechanism the cooperation would not be evolutionarily stable and without the cooperation the mechanism would have no (adaptive) explanation. In Section 3.6 I list many structural features of the biological world that we suggest might plausibly have evolved as social niche modifiers in a process something like that sketched out here (e.g. the self/nonself discrimination mechanisms in filamentous fungi, the apical meristem topology of vascular plants and the obligate co-dispersal of endophytic fungi with their symbiotic sedges).

A number of authors have previously suggested, in general terms, that some sort of runaway social selection between population structure and social behaviour must be at work when we

see the evolution of cooperation in nature (Breden and Wade, 1991; Crespi, 2004; Thompson, 2005; Santos et al., 2006a; Rosas, 2010; Van Dyken and Wade, 2012; Clarke, 2014; Sober and Wilson, 1998, p.97). Michod and Roze (1999) investigated the interplay of social behaviour with a modifier locus (that either imposed a bottleneck or policing) but they built collective-level selection into their model as one of its assumptions (I want to explain how collective-level selection gets started). There have also been a number of more game-theoretic studies investigating the effects of allowing the underlying game, usually a Prisoners' Dilemma, to be changed by the players. Some involve individuals modifying the payoff matrix directly (Worden and Levin, 2007), effectively modifying the payoff matrix by introducing side payments (Akçay and Roughgarden, 2011), or modifying the payoff matrix by modifying assortment by enabling adaptive linking in a network setting (Cao et al., 2011; Pacheco et al., 2006a,b). See also Skyrms (2004).

I should note that social niche modifiers making the conditions for cooperation less favourable could also exist but are of less interest to my project, because they do not have the same potential to drive upward changes in the level of selection and individuality. For a general mathematical treatment of social niche construction exploring the full range of theoretical possibilities, see Jackson (2015); Jackson and Watson (2016). Note that Jackson uses the term 'game changing trait' where I use 'social niche modifier'.

The details of social niche construction vary according to whether the social trait and social niche modifier trait are always inherited together or can be separated by sexual recombination. The former case yields a haplotype-selection regime, while the latter yields an allele-selection regime with epistasis (Neher and Shraiman, 2009). If the population is asexual, such that the genetic basis of the social trait and social niche modifier trait are always passed on together, then it is straightforward to see how a self-reinforcing process like the one adumbrated above could arise. Ever-increasing levels of social niche modification in an asexual lineage lead to ever increasing levels of cooperation in that lineage, leading to greater fitness of that lineage. The result is that it can invade the population by competitively excluding rival lineages unable to avail of the fitness benefits of cooperating (due to their remaining in the original social niche in which cooperation is not possible). If the population is sexual and the social trait and social niche modifier trait do not always co-segregate, then the story is similar but slightly more complicated. Recombination will shuffle the alleles and disrupt advantageous pairings. However, selection will favour certain pairings over others, causing correlations to accumulate in the gene pool (linkage disequilibrium; see also Silver and Di Paolo, 2006). Notably, it will favour a pairing of a cooperation-enhancing social niche modifier with cooperation at the social trait while disfavouring a pairing of the wild-type social niche modifier with cooperation at the social trait. This is an example of selection in the presence of epistasis, a major source of linkage disequilibrium in nature (Kirby et al., 1995). This can mean that, even in the presence of sexual recombination, the pairing of social trait and social niche modifying trait can occur more frequently than would be expected based on their independent frequencies in the gene pool.

3.5.1 Social niche construction and the Extended Evolutionary Synthesis

Mainstream evolutionary biology ('the Modern Synthesis') assumes biological individuality as a given and then asks questions about the circumstances in which individuals will be subject to evolutionary change. It does not generally try to explain where the individuals came from, nor does it allow for individuality itself to be changed by the evolutionary process. Social niche construction theory tries to address these gaps. Due to its emphasis on the role of constructive processes in evolution and on reciprocal causation, social niche construction theory can be understood as belonging to the emerging body of ideas known as the Extended Evolutionary Synthesis (Pigliucci, 2007; Laland et al., 2015).

Note that social niche construction is distinct from 'cultural niche construction', a process of interest in the study of human evolution, whereby human cultural traits can modify the strength and direction of selection on human genes (e.g. Laland et al., 2001; Borenstein et al., 2006). Recall that the 'social trait' in social niche construction is defined as a behaviour that evolves because it affects the fitness of others in addition to the actor (e.g. altruism, mutual cooperation or selfishness) and is not to be confused with the social, societal or cultural traits of interest in the study of human evolution. In this thesis I focus on cases of social niche construction where the social trait and the social niche modifying trait are both genetic (and both vertically transmitted) and I do not discuss any human examples. However, Powers (pers. comm.) points out that some cases of human cultural niche construction might also be cases of social niche construction (e.g. Powers and Lehmann, 2013).

3.6 Many structural features of the biological world can be understood as social niche modifying traits (SNMs)

> *"When we examine a complex social group [i.e. collective life-form] we frequently see, like tourists watching a ceremonial changing of the guard, features that make sense only as the products of a more turbulent past." (Bourke, 2011, p.194)*

Biological examples of social niche modifiying traits include old favourites such as life-history bottlenecks, early-segregating slowly-dividing germlines and worker-policing in social insects. However, rather than special cases or rare curiosities, I follow Bourke (2011) in interpreting a very wide range of structural features of the biological world as playing the functional role of social niche modifiers, ubiquitous wherever complex adaptation is in evidence. In a multilevel setting, social niche modifiers change the effective game being played between particles, moving particle and collective fitness interests into greater alignment.

As I said in the opening paragraph of this chapter, what is new here is not the idea *that* collective life-forms have mechanisms that ameliorate or avoid internal conflict (Michod and Herron, 2006;

Queller and Strassmann, 2009) but to view the *evolution* of such mechanisms as admitting of a general theoretical treatment.

Table 3.1: Examples of structural feature of biological world that function as social niche modifiers for populations of entities interacting in their presence. The first five examples are implicated in egalitarian transitions (unlike-kinds coming together). The other examples are implicated in fraternal transitions (like-kinds coming together).

Structural feature of biological world	**Role as a social niche modifier**
Suppression of segregation distorters in diploids (Maynard Smith, 1958; Leigh, 1971, 1991)	Yields fair meiosis, which avoids intragenomic conflict by placing the alleles at each locus on a diploid genome 'in the same boat' with regard to their chances of reproductive success right up until the moment segregation occurs (Haig and Grafen, 1991). Gene and genome fitness interests are aligned in the presence of fair meiosis and not aligned without it.
Obligate co-dispersal of mitochondria and chloroplast in eukatyotic cells.	Vertical transmission means both partners meet a shared reproductive fate. This aligns the fitness interests of both parties in the symbiosis (Bergstrom et al., 2003)
Obligate co-dispersal of mycetocyte bacteria (operating in the gut) with their insect hosts. In many species, including cockroaches, transmission occurs in the ovaries (Douglas, 1989)	Vertical transmission means both partners meet a shared reproductive fate. This aligns the fitness interests of both parties in the symbiosis (Ewald, 1987).
Obligate co-dispersal of endophytic fungi with their symbiotic grasses and sedges (Clay, 1990).	Vertical transmission means both partners meet a shared reproductive fate. This aligns the fitness interests of both parties in the symbiosis.

Uniparental inheritance of mitochondrial DNA (Birky, 1995, p.149)	Avoids conflict that might occur if there were cytoplasmic chimerism in eukaryote cells (Burt and Trivers, 2006, p.149); also causes nuclear-cytoplasmic conflict over sex ratio (Schnable and Wise, 1998)
Unicellular life-history bottlenecks (Dawkins, 1982)	Alignment of cellular fitness interests in multicellular organisms, due to their clonal relatedness (Dawkins, 1982)
Germline sequestration in metazoans (Buss, 1987)	Denies heritability to selfish cell lineages arising in the soma. The inclusive fitness interests of somatic cells are then best served by supporting the reproduction of the germline cells, rather than attempting to reproduce directly (Michod, 2006; Bourke, 2011)
Apical meristem topology in vascular plants (Klekowski, 1988)	Denies heritability to selfish cell lineages arising outside the apical initials. The inclusive fitness interests of somatic cells are then best served by supporting the reproduction of the apical cells, rather than attempting to reproduce directly.
Allorecognition mechanisms in benthic tunicates (Grosberg, 1988)	Avoids threat of parasitism (free-riding on club goods) that would be present if genetically unlike colonies merged freely.
Allorecognition mechanisms in anenomes (Ayre and Grosberg, 2005)	Avoids threat of parasitism (free-riding on club goods) that would be present if genetically unlike colonies merged freely.
Self/nonself discrimination in filamentous fungi (Glass et al., 2000)	Avoids threat of parasitism (free-riding on club goods) that would be present if genetically unlike colonies merged freely.
Cell-cycle synchronization in myxomycetes (Buss, 1987, p.130)	Turns potentially defector mutations (that increase cell fitness while decreasing plasmodium fitness) into ordinary deleterious mutations (that decrease both cell fitness and plasmodium fitness)

Kin-recognition mechanisms in cellular slime moulds such as Dictyostelium discoideum (Mehdiabadi et al., 2006)	Avoids threat of free-riding that would be present if genetically unlike cells merged freely. High relatedness during the aggregation phase of the lifecycle enables cooperative division of labour between stalk and fruiting-body roles, both of which are necessary for successful reproduction (Bourke, 2011).
Mechanisms of policing, punishment and coercion in eusocial insect societies (e.g. Ratnieks, 1988; Wenseleers et al., 2004)	In the presence of these social niche modifiers, the inclusive fitness interests of workers are best served by supporting the reproduction of the colony (through the queen), rather than attempting to reproduce directly. Policing and punishment modify social niche without modifying assortment.
Copy number control in non-conjugative plasmids	Non-conjugative plasmids are vertically transmitted symbionts of bacteria (Rankin et al., 2011). Intracellular selection favours plasmids that reproduce as fast as possible. But bacterial cells with an optimal number of plasmids are fitter than those with too many (or too few) (Diaz Ricci and Hernández, 2000; Harrison et al., 2012). The social niche experienced by plasmids might be a Prisoners' Dilemma, were it not for plasmid-driven copy-number-control that limits the potential for within-group selfishness (Paulsson, 2002; Kentzoglanakis et al., 2013) and aligns the fitness interests of plasmids within bacterial cells.

Competition for scarce resources in demes of red grouse (Wynne-Edwards, 1962).	No social niche modifier is mentioned in this example. Selfish behaviour is the default case, in no need of special explanation. Even though it might be possible to raise the carrying capacity if individuals exercised consumption restraint, this does not happen. This is because the fitness cost of such restraint is borne fully by the individual exercising it, while the benefit arising from it is enjoyed by the whole group (i.e. a Tragedy of the Commons (Hardin, 1968)). The effective game being played in the social niche is a Prisoners' Dilemma. The unmitigated conflict of interests means cooperation (in the form of reproductive restraint) is not a fit strategy in such a social niche.

3.7 A broad classification of social niches (collective action involving like- and unlike-kinds)

Collective action can be divided into two broad categories, the first involving the coming together of like kinds and the second involving the coming together of unlike kinds - in both cases to enjoy some mutual benefit that could not be accessed through solitary action (Mill, 1848, Book I, Chapter VIII). The first case includes colonial organisms and can sometimes lead to 'fraternal' evolutionary transitions. The latter case includes inter-specific symbioses and can sometimes lead to 'egalitarian' transitions (Queller, 1997). Social niche modifying traits also fall into two corresponding categories.

The first category involves like-kinds coalescing into higher-level units that, in extreme cases (fraternal transitions) come to be evolutionary units in their own right. Examples include colonial organisms of many kinds. These cases rely on *genetic relatedness* between the coalescing entities to align fitness interests. The pathway to the fraternal transitions can be understood in terms of the evolution of social niche modifiers affecting the relatedness between social partners (e.g. kin recognition or population viscosity). Relatedness is a very common realizer of the assortment of social behaviours.

The second category involves unlike-kinds coming together. It includes inter-specific symbioses and can sometimes lead to 'egalitarian' transitions. Egalitarian transitions rely on both parties to the coalition retaining their ability to reproduce, albeit within a mechanism that ensures they each do so only if the other does also (Queller, 1997). This forced *shared reproductive fate* aligns the fitness interests of both parties (Ewald, 1987; Bourke, 2011). The pathway to egalitarian transitions can be understood in terms of the evolution of social niche modifiers affecting

the obligate co-dispersal of both partners (sometimes called 'vertical transmission'). Other social niche modifiers, such as partner-discrimination and sanctioning may also be important for maintaining the positive assortment of cooperative behaviour in inter-specific mutualisms with horizontal transmission (Noë, 2001; Sachs et al., 2004; Kiers et al., 2011; Archetti et al., 2011). In both the fraternal and egalitarian cases, while the unmodified social niche may feature conflict between particle and collective interests, the modified social niche finds those fitness interests brought into alignment.

3.8 Many social niches are possible

To understand the evolution of social niche modifiers, it is necessary to understand the social niches they are modifying. Social niches that are 'red in tooth and claw', involving straightforward competition or predation are usually taken to be the default case and require no special explanation (pace Roughgarden, 2009). In these cases, selfish behaviour is evolutionarily stable. An individual can gain only at the expense of another and there is no opportunity to increase overall social welfare.

Things start to get interesting when we turn to the social niches that represent a cooperative dilemma (Dawes, 1980). Here, there is an opportunity to avail of new fitness benefits available only through collective action (Olson, 1965; Calcott, 2008). However, such collective action is often undermined by cooperative dilemmas (Olson, 1965; Maynard Smith, 1988). Individually rational behaviour leads to outcomes that are not the best possible outcome for the individuals concerned (Macy and Flache, 2002). In cases such as this, we can usefully employ concepts from game theory to describe, categorise and explain the properties of a social niche. It is also in these cases that social niche construction has something new to offer because it can dissolve the apparent paradox that usually blocks attempts to explain, in terms of individual-level selection, how group-beneficial outcomes can arise in spite of the presence of a cooperative dilemma.

3.8.1 Analysis of collective action problems between like-kinds using the T-S plane (synchronic)

In the remainder of this chapter, I present a minimally technical analysis of social niche construction among like-kinds. The like-kinds case has the advantage of greater simplicity of exposition and tractability of analysis, as one can model such collective action problems as the linear aggregate of payoffs in a pairwise two-player two-strategy symmetric game[5] (Hamilton, 1975). However, I emphasise that my general claims about social niche construction (above) are intended to be understood more broadly than the detailed treatment for cooperation among like-kinds (including fraternal transitions) presented in the remainder of this chapter.

[5] See Archetti (2009) and Archetti and Scheuring (2012) for discussion of some more biologically realistic alternatives to two-player symmetric games - all of which are more rather than less amenable to cooperation.

The games I use to model social niches experienced by groups of like-kinds involve two fungible players, each with two strategies that we tentatively[6] label 'C' and 'D'. Thus there are four possible payoffs to a focal individual (payoff matrix in Table 3.2). All possible types of social

TABLE 3.2: Following convention (Axelrod and Hamilton, 1981), let R be the payoff for mutual 'C', S for unilateral 'C', T for unilateral 'D' and let P for mutual 'D'. Payoffs to row player are shown.

	C	**D**
C	R	S
D	T	P

dilemma in a two-player two-strategy symmetric game can be described by different orderings of R, S, T and P (Macy and Flache, 2002):

- Prisoners' Dilemma $(T > R; P > S)$
- Stag Hunt $(R > T; P > S)$
- Snowdrift $(T > R; S > P)$
- Harmony Game $(R > T; S > P)$

Santos et al. (2006b) have introduced a compressed representation of the space of all such dilemmas on a single two-dimensional space[7] that I will refer to as the 'T-S plane' (Figure 3.1). This is a very useful tool for thinking about social evolution. The Prisoners' Dilemma, Stag Hunt, Snowdrift and Harmony Game are often considered separately but are in fact continuous with one another. Synchronically, any[8] social niche can be characterised by a single point on the T-S plane. Diachronically, social niche construction involves movement across the T-S plane. Figure 3.2 provides an example from the natural history of a colonial marine invertebrate.

Recall that the term 'social niche' refers to the circumstances dictating the strength and direction of selection on social behaviour (3.5). As Jackson (2015) points out, this has a simple representation on the T-S plane. Consider the point in the centre of Figure 3.1, at $T = 1, S = 0$. At this point, the payoff matrix $\left(\begin{smallmatrix} R & S \\ T & P \end{smallmatrix}\right)$ is $\left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$. The player's own strategy is not a difference-maker to its payoff (and hence fitness), so there will be no selection on social behaviour at this point. Call this the origin. Then, for any point on the T-S plane (i.e. for any social niche in scope here):

- The strength of section on social behaviour is proportional to the distance from this point.
- The direction of selection on social behaviour is given by the difference between the actual frequency of co-operation in the population and the equilibrium frequency of cooperation

[6] Whether playing 'C' can be properly described as 'cooperation', 'strong altruism' or something else depends on further details of the game being played - see Section 3.8.3 below.

[7] See also Weibull (1995, p.29).

[8] Subject to the stated assumption that the social niche can be modelled by aggregating a two-player two-strategy symmetric game.

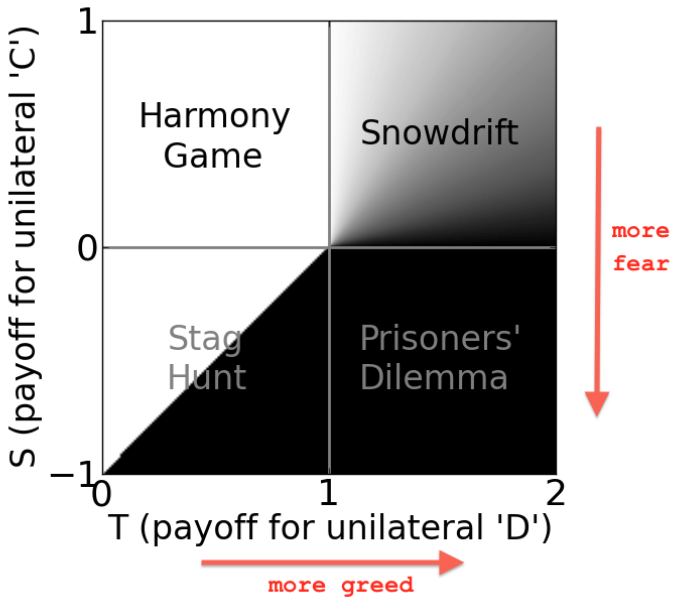

FIGURE 3.1: **The T-S plane**: Conflict between individual and collective interests can be modelled as the aggregate outcome of pairwise rounds of a two-player symmetric game between individuals. The game, its dynamics and its equilibria are the determining characteristics of a social niche, as they (in conjunction with the frequencies of strategies in the population) determine the strength and direction of selection on social behaviour. The payoff matrices of these games are conventionally represented as a 4-tuple of the four possible payoffs (R, S, T, P), listed in Table 3.2. The Stag Hunt, Prisoners' Dilemma, Snowdrift and Harmony games are often considered separately. However, by normalising the payoff matrix so that $R = 1$ and $P = 0$ and limiting $(0 \leq T \leq 2)$ and $(-1 \leq S \leq 1)$, the space of all such games can be represented on a single continuous 2D plane with dimensions S (the payoff for unilaterally playing 'C') and T (the payoff for unilaterally playing 'D') (Santos et al. 2006b). Shading indicates equilibrium level of cooperation (black=0). **Synchronically, any* social niche can be characterised by a point on the T-S plane. Diachronically, social niche construction involves movement across the T-S plane.** The evolution of social niches supporting cooperation requires that initially conflicted social niches be translated into instances of the Harmony Game or Snowdrift game (for full or partial cooperation respectively). Diagram adapted from Santos et al. (2006b, Figure 2). * Any, subject to the restriction that it can be represented with a two-player two-strategy symmetric game between like-kinds.

at that point (as represented by the shading in Figure 3.1). (Picture polar coordinates with a radius of some length at some angle θ.

Biological examples of social niches instantiating Prisoners' Dilemma, Snowdrift and Stag Hunt games

Each of the two-player two-strategy symmetric games is instantiated in the biological world in many different contexts. (Not all of them are equally implicated in major evolutionary transitions but I will not pre-judge that issue here.)

The Prisoners' Dilemma is a very common social niche. For example, many fishes in the sea bass family (Serranidae) are hermaphrodite. Individuals produce both eggs and sperm but are not self-fertile. Due to anisogamy, egg production is more costly than sperm production. When pairs mate, both individuals benefit from the efforts of the egg producer but the cost is paid unilateral by the individual in the female role (Connor, 1992). An individual adopting the male role more than half the time would thus gain an advantage ($T > R$) but if all individuals adopted the male role then all would suffer ($P < R$). The social niche is thus a Prisoners' Dilemma.

Social niches involving a Snowdrift game are also common in nature. Brewer's yeast (Saccharomyces cerevisiae) is a single-celled eukaryote that breaks down sucrose from a food source by secreting the enzyme invertase. Invertate is costly to produce and, crucially, hydrolises sucrose *outside* the cell wall where it is effectively a club good. Should a particular mutation occur at the SUC2 locus, it is possible for the production of invertase to be suppressed. Gore et al. (2009) found that populations consisting of high densities of the mutant type could be invaded by the wild type, while populations consisting of high densities of the wild type could be invaded by the mutant type. Polymorphic mixtures of the types were found to be stable at a certain ratio of densities. This is a clear case of the dynamics expected in a Snowdrift game. (See also Greig and Travisano, 2004).

Stag Hunt games are also common. A strain of the bacterium E. coli produces an antibiotic called colicin that destroys rival strains. Generation of the toxin incurs a fitness cost to the individual producing it. However, it also provides a benefit to the bearer (greater than the cost), as long as many others also produce the antibiotic. When the colicinogenic strain is common in freely mixed populations, the rival wild-type strain is suppressed and the colicinogenic strain prevails. However, when the colicinogenic strain is in the minority, the costs for each individual of producing the colicin outweigh the small suppressant effect it has on the rival strain. So, while the colicinogenic strain is stable against invasion by the wild type, it cannot invade the wild type when rare. This is, therefore, a social niche featuring a Stag Hunt game (Chao and Levin, 1981). If the experiment is repeated on an agar plate instead of in a freely mixed liquid environment, the positive assortment of social behaviours brought by the spatial proximity of relatives allows the colicinogenic strain to invade a population of wild-type E. coli even when founded by just a few colicinogenic cells (Skyrms, 2004). For an example involving a Stag Hunt game in biofilm formation in Pseudomonas fluorescens, see Rainey and Rainey (2003).

Finally, a Harmony Game is representative of the social niche experienced by the cells in a multi-cellular organism where all cells are clonally related to one other.

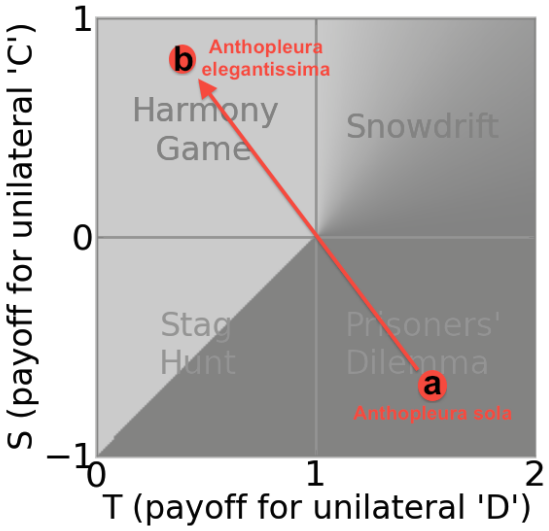

FIGURE 3.2: I offer a putative example of social niche construction in sea anenomes. **(a) Anthopleura sola** is a solitary-living sea anenome. Individuals compete for space on the rocky shore. It employs sexual reproduction. Offspring disperse after reproduction, so any neighbouring individuals in adjacent areas are unlikely to be close relatives. If individuals exercised growth restraint, this would allow a larger absolute biomass of Anthopleura sola because smaller individuals are more efficient at converting food to self. However, the cost of exercising reproductive restraint would be visited entirely upon the individual exercising it, while the benefit would be enjoyed by the whole community. They thus live in a social niche characterised by a Prisoners' Dilemma in which the pro-social strategy is individually maladaptive and not evolutionarily stable. **(b) Anthopleura elegantissima** is a colonial anenome that shares a solitary-living common ancestor with Anthopleura sola (Francis, 1979; McFadden et al., 1997). Anthopleura elegantissima colonies grow vegetatively on the benthic substrate, such that adjacent polyps are clonaly related. This means individual and colony inclusive fitness interests are aligned. Colonies share club goods within the colony and antagonism in Anthopleura elegantissima is between colonies rather than between polyps (Ayre and Grosberg, 2005). The phylogenetic tree for the Anthopleura is complex with clonality arising, being lost and arising again numerous times (Geller and Walton, 2001). I tentatively suggest the changes to the social niche experienced by the polyps can be understood in terms of social niche construction, where the social niche modifying trait is one that modifies the life-history of polyps, particularly their propensity for limited dispersal after vegetative reproduction (Geller and Walton, 2001).

3.8.2 Social niche construction as movement on the T-S plane (diachronic)

Social niche construction was described in general terms in Section 3.5 above. Turning back to Figure 3.1, social niche construction can be represented as movement of the social niche across the T-S plane. The ways it can move depend on which social niche modifier is acting; different social niche modifiers act in different ways. For example, punishment reduces the gains to be made from unilateral defection, thus reducing T while leaving S unchanged. This can be represented as a shift to the left on the T-S plane. The effect of increasing assortment of strategies (perhaps by increasing genetic relatedness among interaction partners) is described fully in the next section (Section 3.8.3). As a quick preview: assortment moves any initial social niche in a straight line on the T-S plane toward the Harmony Game (point $S = 1, T = 0$), representing increasing toward alignment of individual and collective interests.

Consider the shading in Figure 3.1. Where the initial social niche lies in a region where there is a gradient in the equilibrium level of cooperation (i.e. the Snowdrift region only), small changes in social niche immediately yield small changes in the equilibrium level of cooperation and the social niche construction process can proceed by gradual changes to the social niche. Where the initial social niche lies in a region with no gradient in the equilibrium level of cooperation (e.g. the Prisoners' Dilemma region and the defection-dominant region of the Stag Hunt), small changes in social niche will not immediately alter the equilibrium level of cooperation. In these cases, a larger translation on the T-S plane is required. For example, consider the evolution of multi-cellularity, which has evolved independently many times (Grosberg and Strathmann, 2007). In each of the land plants, the red algae, the brown algae, the animals and the fungi, the transition to multicellularity involved the evolution of a life-history involving cells 'staying together' after division rather than dispersing (Fisher et al., 2013). Starting among a wild type that disperses after cell division, a mutant social niche modifier elevating the probability of adhesion after cell division would, in those cases where adhesion did occur, confer very high levels of trait assortment when compared to the freely-mixing wild type that dispersed after cell division. One need not assume this is a silver-bullet mutation that radically changes the social niche in a single generation - an evolutionary process that slightly elevated the probability of adhesion after cell division could effect such a change gradually, leading to higher and higher levels of trait assortment among the mutant strain, ultimately leading to its invasion.

3.8.3 Social trait assortment is a common and powerful social niche modifier

There are many potential social niche modifiers. In this section I focus on one that gets a lot of attention - the assortment of behaviours or traits. In the literature on the the evolution of altruism[9], it is widely held that positive assortment of cooperative behaviour is the key ingredient. Godfrey-Smith (2009, pp.118-120) gives a useful review, summarising that "the familiar mechanisms behind the evolution of altruism can be seen as different ways of achieving correlation between the traits or behaviours exhibited in a population - a tendency for like to accompany like". (Hamilton, 1975; Eshel and Cavalli-Sforza, 1982; Michod and Sanderson, 1985; Sober, 1992; Skyrms, 1994; Godfrey-Smith, 2009; Fletcher and Doebeli, 2009). Strategy assortment - the tendency for like strategies to accompany like - is indeed a powerful social niche modifier, biologically instantiated in a number of ways, including relatedness. But while assortment of strategies is sufficient for pro-social behaviour to be stable in a Prisoners' Dilemma, this is not true of all two-player two-strategy symmetric games. There are other games and other social niche modifiers.

For games where $S < P$ (the Stag Hunt and Prisoners' Dilemma), there is no direct benefit to the actor from its own unilateral playing of 'C', so 'C' can only be an evolutionarily stable strategy if it is reliably reciprocated. In the Prisoners' Dilemma this means that for 'C' to be

[9] 'The evolution of altruism' and 'the evolution of cooperation' are sometimes distinguished in a principled way (not always consistently) and sometimes run together. Which term is correct depends on the underlying game implicit in each case. A full account is given below, in this section.

ESS, there must be circumstances that cause there to be assortment of cooperation. (Recall that I restrict discussion to assortment as a social niche modifier in this subsection. Other social niche modifiers such as side-payments or punishment are not in scope here). In the Stag Hunt, if the balance between the payoffs and the frequency of cooperators in the population dictates that defection is favoured by selection, then for cooperation to evolve from those conditions there must be assortment of cooperation (Skyrms, 2004, Postscript). These are the cases where assortment of cooperative behaviour is necessary (and where strategy 'C' might reasonably be called altruism). In Stag Hunts where the standing frequency of cooperators in the population is sufficiently high, it may be worth chancing it even without any mechanisms causing positive assortment. (In these cases 'C' might reasonably be called mutual cooperation).

There is another class of games: those where the benefit accruing to an actor playing 'C' exceeds the cost of creating it (i.e games where $S > P$). This class includes all Snowdrift (Hawk-Dove) games. Pepper's (2000) example of games involving whole-group beneficial traits that provide more benefit to the actor than they cost (even though the benefit is shared with others) are in this category. In such social niches, positive assortment of cooperation is not necessary to make the 'C' trait evolutionarily stable at some non-zero level. 'C' will be viable at some non-zero equilibrium level (called the 'mixed-strategy equilbrium') in virtue of it providing direct benefit to the actor, even in the absence of positive assortment. (In these cases, 'C' might reasonably be called mutual cooperation.) While positive assortment is not necessary in these cases, it is sufficient if it does obtain: any external factor increasing positive assortment in a Snowdrift game (where $S + T < 2R$) will increase the level of cooperation and collective welfare. In the case of Snowdrift games where $S + T > 2R$, it is not the case that collective welfare increases with positive assortment of strategies, or with greater frequency of the 'C' strategy. In these interesting cases, division of labour between the two strategies produces the best collective payoff and increasing positive assortment can be detrimental (Tudge et al., 2016). It is inappropriate to call the 'C' strategy cooperation in such cases.

Pepper's (2000) 'whole-group traits' provide another way to think about cases where positive assortment is not always necessary. Suppose the total benefit generated by the actor is b, the cost to the actor is c and there are n beneficiaries, each of whom (including the actor) receive b/n. When $b/n < c$, the social niche instantiates a Prisoners' Dilemma and positive assortment of 'C' would be required for it to be evolutionarily stable. However, when $b/n > c$, the game instantiated is in fact a Snowdrift in which 'C' will be stable at some non-zero level *even in the absence of positive assortment of strategies*. So a social niche modifier that reduced n (some sort of preference for small groups, perhaps) while leaving b and c constant could modify a Prisoners' Dilemma into a Snowdrift without directly modifying the assortment of strategies.

In group selection models of the kind discussed by Wilson, D.S. (1975) and Sober and Wilson (1998), assortment can arise in two ways. It can arise at the group-formation stage or it can arise due to particles reproducing within groups. When particles reproduce within groups, assortment can arise if offspring resemble their parents (as they are usually assumed to do in such models, rare mutations excepted). In models that assume only a single round of game-play within groups

before group dispersal and re-formation, this second source of assortment is precluded, often without explicit declaration (and perhaps unconsciously). See Fletcher and Zwick (2004) for a discussion of this point. It is relevant to modelling work in Chapter 4 of this thesis. What matters is social trait assortment, not its source. Social trait assortment due to assortative group formation and social trait assortment due to vegetative reproduction with limited dispersal (and social trait assortment due to anything else) all amount to the same effect when plotted on the T-S plane.

See also Okasha (2006, Section 6.6) for a survey of some of the literature on the relation between assortment, altruism (in that text called 'strong altruism) and cooperation (in that text called 'weak altruism') from the 1980s and 1990s. It seems likely that, had the authors made their assumptions more explicit (perhaps by plotting a point on the T-S plane to illustrate the social dilemma they were discussing), much talking-past one another could have been avoided.

3.8.4 Social trait assortment (an SNM) acts as a translation on the T-S plane

Modifying a social niche by modifying assortment has a quantifiable effect on the effective game being played, with a geometric interpretation on the T-S plane (see Figure 3.3). To see this, consider the following reasoning due to Adam Jackson (Jackson, 2015; Jackson and Watson, 2016) after Grafen (1979). Let degree of assortment α mean that with probability α an individual will play another with the same social trait as itself while with probability $(1-\alpha)$ it will play a randomly selected (non-self) member of the population. When $\alpha = 1$, the game being played is the Harmony Game $\left(\begin{smallmatrix} R & R \\ P & P \end{smallmatrix}\right)$, when $\alpha = 0$, the game being played is the unmodified underlying game $\left(\begin{smallmatrix} R & S \\ T & P \end{smallmatrix}\right)$ and when α is between 0 and 1 the effective game is $\left(\begin{smallmatrix} R & S+\alpha(R-S) \\ T+\alpha(P-T) & P \end{smallmatrix}\right)$. This has the neat geometric analogue on the T-S plane of assortment interpolating along the straight line between the point (T, S), the base game, when there is no assortment to the point $(0, 1)$, the Harmony Game, when there is full assortment. In summary, an underlying game plus some assortment is formally equivalent to another game in a freely-mixed population[10].

[10] A similar result for relatedness (an instantiator of social trait assortment) is given by Taylor and Nowak (2007).

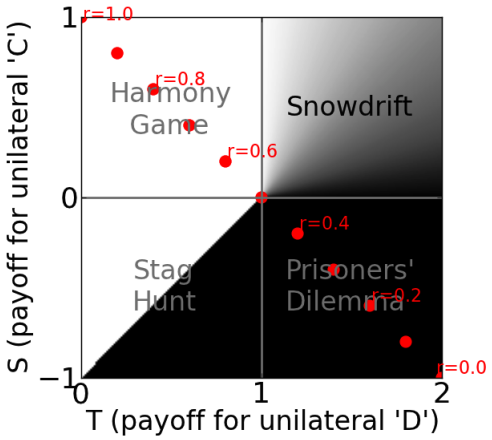

FIGURE 3.3: **Relatedness as a social niche modifier:** Social niche modifiers alter the direction and strength of selection on social behaviour. Between like-kinds, genetic relatedness is usually the most important social niche modifier. (Between unlike-kinds it is typically something else, such as a life-history involving co-dispersal of mutualists). The figure shows the effect of relatedness, r, as a social niche modifier among conspecifics. When games are played between relatives, the expected payoffs are modified due to the elevated possibility that the interaction partner may play the same strategy as the focal individual (Grafen, 1979). In a social niche that between non-kin ($r = 0$) would be a Prisoners' Dilemma, the effective game being played between diploid full sibs ($r = 1/2$) is less conflicted and a game played between clones ($r = 1$) is not conflicted at all. Relatedness between interaction partners is a promoter of the assortment of social traits.

3.9 Social niche construction can explain evolutionary transitions in individuality

In this section, I show the significance of SNC for explaining evolutionary transitions in individuality. I use Sober's (1984, p.97) *selection-of/selection-for* distinction to argue that social niche modifying traits selected-for the alignment of fitness interests of particles within collectives can concomitantly effect *selection-of* a change in the level at which the Darwinian machine operates. That was the thing to be explained.

Social niche construction can explain evolutionary transitions in individuality, as I have defined them. I started this thesis by specifying an explanatory goal (Section 1.4). I set myself the task of explaining how an evolutionary process acting at particle level in a two-level hierarchy could cause a transition such that the evolutionary process could come to act at collective level rather than particle level (and where the evolutionary process is to be understood in terms of character-fitness covariance). The problem appeared to be difficult because it is not easy to see how could the evolutionary process at collective-level ever get started, given that it seems to be faced with a bootstrapping problem.

Social niche construction explains the advent of a collective-level evolutionary process as a side-effect of a particle-level evolutionary process that aligns the fitness interests of particles within collectives as a means to making those particles fitter. To see how, consider Price's covariance

formalism. Evolutionary change occurs in a population of entities when there is a non-zero covariance between those entities' fitnesses and the character of their offspring. We can break this down into a pair of conditions[11], one about selection and one about heredity. Evolutionary change occurs in a population of entities when there is:

1. covariance between entity fitness and entity character (selection)
2. covariance between entity character and mean offspring character (heritability)

Before an evolutionary transition, particles meet conditions 1 and 2 by definition. Before an evolutionary transition, collectives fail condition 1 because there is no covariance between their fitnesses and their characters. Before an evolutionary transition, collectives fail condition 2 because within-collective selection denies collective-level heritability. This is because within-collective selection changes the distribution of particle types over the lifetime of a collective, such that the particles it consists of at maturity (and that will go into its propagules, howsoever formed) will not have the same type distribution as those from which it was itself founded.

After an evolutionary transition, collectives meet conditions 1 and 2 by definition. Particles fail condition 1 (they do not vary in fitness within their collectives, so there can be no co-variance) but still meet condition 2 (there is still particle-level heritability). As we saw in Section 3.5, social niche construction can be understood in terms of the evolution of cooperation among particles, so aligning within-collective fitness interests. But if we think of that same process in terms of 1) selection and 2) heritability at both levels, we can now see that, for particles, the alignment of fitness interests within collectives is also the reduction of fitness variance within collectives (and hence any character-fitness covariance) and thus the negation of condition 1 for particles. Alignment of fitness interests within collectives means all particles in a collective get the same fitness. Simultaneously, for collectives, that same process involves an increase in character-fitness covariance among collectives. The character here is the conflict suppression mechanisms - the social niche modifier and fitness is a function of the extent to which the constituent particles cooperate. Finally, that same social niche construction process also increases the extent to which collective-level heredity obtains (condition 2), by suppressing within-collective change (Frank, 2012).

So while there is no selection-for collectives being the unit of evolution, there is selection-of the conditions for them to be so. The selection-of/selection-for distinction is as per Sober (1984). The oblique way in which early-stage evolutionary transitions are explained by social niche construction thus side-steps some of the problems associated with other attempts to explain them - I return to this point in the Discussion (Section 7.5). I offer social niche construction as an explanation for how collective-level selection could get started and thereby how evolutionary

[11] As Okasha (2006, p.37) points out, the move from the first, one-part version of the Price formalism to the two-part version given here is not fully general because covariance is not generally a transitive relation. See McNemar (1962); Sotos et al. (2009). This technicality need not overly concern us here: assume further that the two covariances are sufficiently strong that there *is* covariance between entity fitness and offspring character.

transitions in individuality might be enabled. Of course, once there *is* a collective-level selection regime in force then a collective-level evolutionary process can evolve division of labour, specialisation of parts and all manner of complex adaptations.

3.10 Note: Social niche modifiers are not greenbeards

I pause briefly to note that social niche modifiers are not greenbeards. A greenbeard is a trait that has evolved because it signals to others that its bearer is a cooperator (Gardner and West, 2010). The idea was first suggested by Hamilton (1975). Even in social niches where playing strategy 'C' incurs direct costs for the actor (and is thus strong altruism), greenbeards can stabilise cooperation on the condition that the green beard gene and the altruistic gene remain tightly linked (or pleiotropic expressions of the same gene). However, greenbeards are vulnerable to cheater mutants ('falsebeards') that bear the green beard but play the 'D' strategy (or mutations to the pleiotropic gene, such that it causes green beards but defection) (Gardner and West, 2010).

Social niche modifiers are not vulnerable to cheating in this way because they are not merely signals ('cheap talk' as Skyrms, 2004 calls it) but ways of living that cannot be faked. For example, a trait that signals to others that the bearer interacts preferentially with relatives (while the bearer in fact interacts with random members of the population) would not be social niche modifier in my sense, because it does not cause the consequences of social behaviour to preferentially fall upon similar individuals.

3.11 Comparison with earlier work on social niche construction by Powers

Previous published work on social niche construction (Powers, 2010; Powers et al., 2011)[12] has focused on an MLS1 (Damuth and Heisler, 1988) group-selection model, based on D.S. Wilson's (1975) 'trait group' model for the evolution of cooperation. That work demonstrated the concurrent evolution of cooperation and of a population structure supporting cooperation (i.e. a social niche modifier), despite the presence of a social dilemma. Specifically, it demonstrated the evolution of an SNM that increased return-on-investment to the actor in a Snowdrift game (where benefit is b/n) by reducing group size, n (thus increasing S, the payoff for unilateral cooperation). In common with all MLS1 models, it involved the evolution of a particle-level property (cooperation) in a global population of particles. Collectives in MLS1 models provide a selective context for particle-level selection but are not themselves units of evolutionary change (Maynard Smith, 1987; Okasha, 2006).

[12] I am greatly indebted to Powers (pers. comm., ibid.) for introducing me to the concept of social niche construction.

The move from Powers' work on social niche construction to my current position owes much to a revisiting of Maynard Smith's (1976; 1987) criticism of the trait-group models employed by D.S. Wilson (1975). Maynard Smith argued that these models were not capable of demonstrating true group-level adaptation because their groups lacked group-level heritability. The consequence, he argued, is that these groups do not qualify as 'units of evolution', even though they might be described as 'units of selection' (see also Okasha, 2005a). In this thesis, I am interested in the evolution of social niche modifiers supporting both increased cooperation *and* increased heritable fitness variance at collective level, such that collectives increasingly become 'units of evolution'.

Chapter 4

The evolution of the life-history bottleneck as a social niche modifier

4.1 Overview of chapter

A life-history bottleneck involves a collective life-form consisting of many 'particles' reproducing by means of a *unitary propagule* - a stage in the life cycle that consists of only one particle. It is a very common life-history trait that has evolved many times independently in diverse taxa, including the animals, land plants, fungi, brown algae and red algae and many others. I start

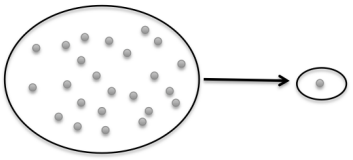

FIGURE 4.1: Schematically, a life-history bottleneck can be understood as a reproductive mechanism employed by collective-living lifeforms. New collectives are founded by a single particle which then goes on to multiply, so forming a new collective.

this chapter with a brief survey of the natural history of the bottlenecked life cycle. I then examine the potential for such bottlenecks to be understood as instantiating the role of Social Niche Modifiers, finding that they can cause the assortment of social behaviour in some circumstances. I describe a simple two-locus computational model of the evolution of a life history bottleneck in the life cycle of a population of particles organised into collectives, where the ancestral state features an unbottlenecked life history. The model allows me to test the theory that bottlenecks act as social niche modifiers in a pro-social way. It also allows me, using Clarke's (2010) Price Approach to biological individuality (introduced in Section 2.3.3 above), to study the change in the relative level of character-fitness covariance - and so individuality - from particle-level to collective-level, that accompanies bottleneck evolution.

From this modeling work, I take away a number of things. One is a clearer conviction that social niche construction could be a workable mechanism for the concurrent evolution of both cooperation and the structural circumstances that support it. Another important message is that some cases of the evolution of cooperation are also cases of transitions in the hierarchical level at which character-fitness covariance obtains (and that is the main thing to be explained in this thesis). A further lesson is about the idea that particles somehow sacrifice or give up fitness for the sake of collective good (Michod et al., 2006) - I find this problematic. This will feed into the discussion in chapters 5 and 6 about what happens to fitness during evolutionary transitions in individuality.

4.2 What is a life-history bottleneck?

I start by explaining what a life-history bottleneck is and how it differs from the alternative possible life histories. Perhaps due to their commonness, newcomers to the subject often struggle to recognise that life-history bottlenecks are but one possibility among many and that their widespread distribution across taxa is something in need of explanation, rather than the default position. Life-history bottlenecks are multiply-realizable - they occur in many phyla at many levels in the biological hierarchy, with very different types of things in the roles of particles and collectives in each case.

Wallace on possible life-histories

A collection of English language translations of Weismann's writings was published in 1889 under the title *Essays on Heredity and Kindred Biological Problems*. One of the translators, E.B. Poulton, included the following note which he received from A.R. Wallace as part of Wallace's review of the edition. Wallace is commenting on Weismann's essay 'The Duration of Life', one of the first attempts to offer an evolutionary explanation for senescence but he starts by discussing the space of theoretically possible life-cycles (including indefinite persistence, binary fission and propagule emission).

> "The Action of Natural Selection in Producing Old Age, Decay and Death.
>
> Supposing organisms ever existed that had not the power of natural reproduction, then since the absorptive surface would only increase as the square of the dimensions while the bulk to be nourished and renewed would increase as the cube, there must soon arrive a limit of growth. Now if such an organism did not produce its like, accidental destruction would put an end to the species. Any organism therefore that, by accidental or spontaneous fission, could become two organisms, and thus multiply itself indefinitely without increasing in size beyond the limits most favourable for nourishment and existence, could not be thus exterminated: since the individual only could be accidentally destroyed,–the race would survive. "
>
> - A.R. Wallace, translator's note in the English-language edition of Weismann (1889, p.23).

Multicellularity evolved from single-celled life. In principle, the most simple way to be a multicellular organism (or any collective made of particles) is simply to exist and persist as a single entity without a life cycle. One can imagine a great mass of cells, some dividing, some dying, the whole mass persisting over some period of time. This possibility was considered by Wallace (see boxed text). He saw that such a way of life is inherently unsustainable because, without reproduction and in the face of cumulative exogenous mortality risk, such a life form could not survive indefinitely. I am unaware of any real-world examples of such a species but it is a useful conceptual starting point.

Wallaces' point is that for a species to exist in a stable manner over an extended period of time, reproduction of its members is necessary. With reproduction necessarily comes life cycle (Bonner, 1965). A simple form of life cycle for a collective is fission: the collective splits into two parts, they move apart, each part is now a new collective. This life-cycle (sometimes called fissiparity) is common, occurring in some multi-cellular animals (e.g. Emson and Wilkie, 1980), most if not all mycelial fungi (Burnett, 2003), many plants (e.g. Degennaro and Weller, 1984), colonial cnidarians (Farrant, 1985), colonial arachnids (Aviles, 2000) and protists (Anderson, 2013).

Broadly speaking, there are two ways to be multicellular. The first (the sub-social pathway to multicellularity, 'staying together') is to start from a single cell and remain connected rather than separate after cell division. The second (the semi-social pathway to multicellularity, 'coming together') is to aggregate together from separate cells. Both of these life histories have evolved many times in disparate taxa. Only staying together results in obligate multicellularity (Fisher et al., 2013; West et al., 2015) and complex multicellularity with sophisticated division of labour (Bourke, 2011). It is multicellularity through 'staying together' first which concerns us here.

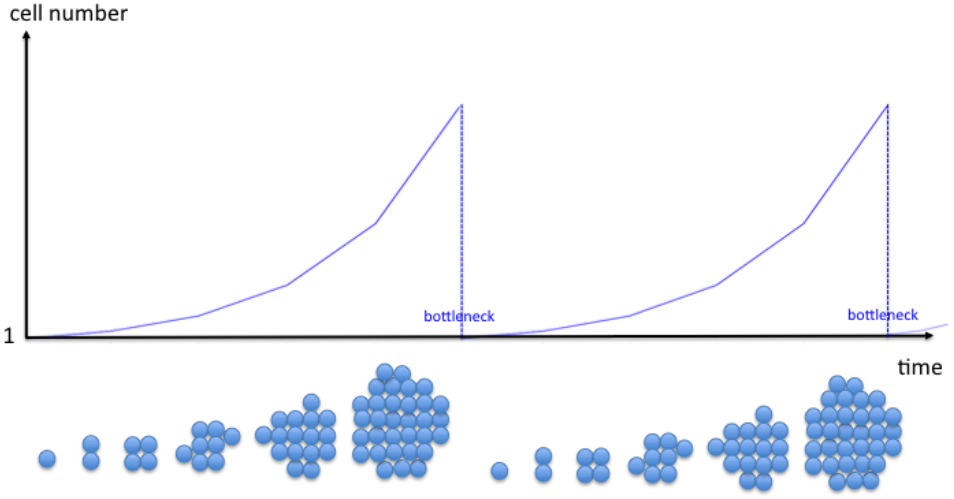

FIGURE 4.2: Schematic plot showing relationship of cell number with time for a bottlenecked life cycle. The point at which a small part of the parent breaks off to become the offspring is the bottleneck. For simplicity, semelparity (one reproduction event per life cycle) with clutch size one is assumed. Each organismic generation starts off with low cell number (one cell in the case of maximal bottlenecking), grows to many cells, reproduces by emitting a single-celled offspring and dies. Compare with Figure 4.3

In this life history, a single cell divides many times (by mitosis) resulting in a multi-cellular organism all derived from and highly related to the same original cell. At a later point in its life cycle it produces offspring as single-celled propagules which repeat the process. This single-celled propagule is termed a 'bottleneck' because it represents a dramatic narrowing in the cell lineage's numbers (Figures 4.2 and 4.3). A unicellular bottleneck of this sort is found in the life-history of very many organisms (see next section for a survey).

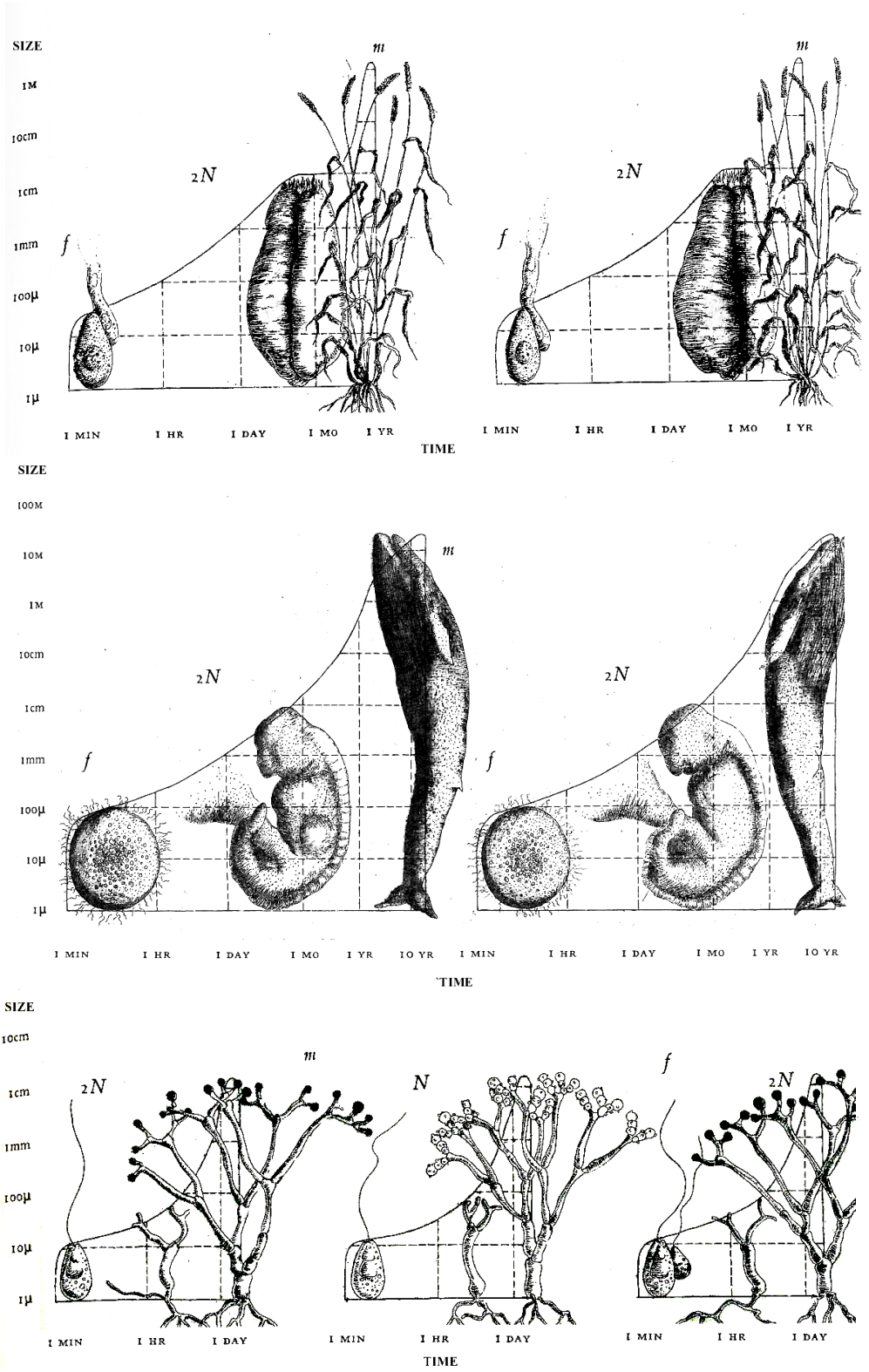

FIGURE 4.3: Illustrations from Bonner's (1965) *Size and Cycle*, showing the bottlenecked life-histories of wheat (a land plant), the blue whale (an animal) and allomyces arbuscula (a fungus). The horizontal axis represents time on a log scale, the vertical axis represents size on a log scale (consider size a proxy for cell number). Compare with Figure 4.2. Reproduced with permission of Princeton University Press.

4.3 Bottlenecks are multiply realised, in diverse taxa and at several hierarchical levels

The purpose of this section is to show that life-history bottlenecks are very common among collective life-forms, being involved in the evolution of many examples of multicellularity, eusociality and coloniality in widely disparate taxa and at many levels of the biological hierarchy. I include some familiar examples and some examples that are not normally interpreted as life-history bottlenecks but, I contend, should be.

4.3.1 All kingdoms of complex multi-cellular eukaryotes employ life-history bottlenecks

I will start with multi-cellularity. Life-history bottlenecks are very common among multicellular phyla (Hamilton, 1964; Bonner, 1965, 1974; Dawkins, 1982; Buss, 1987; Maynard Smith, 1988; Grosberg, 1988; Grosberg and Strathmann, 2007). Multi-cellularity has independently evolved at least twenty-five times in the prokaryotes and eukaryotes (Grosberg and Strathmann, 2007). Many multi-cellular forms are simple, involving one or few cell types and no reproductive division of labour. There are numerous examples from both the prokaryotes and eukaryotes (Grosberg and Strathmann, 2007; Bourke, 2011). While some of these simple multicellular species employ life-history bottlenecks, many others do not (Fisher et al., 2013, Figure 2). When we turn to the instances of complex multi-cellularity (i.e. exhibiting reproductive division of labour and different somatic cell types), a clear pattern emerges (Fisher et al., 2013). Complex multi-cellularity is not found in the prokaryotes (Bourke, 2011). Complex multi-cellularity in the eukaryotes is found only where a single-cell life-history bottleneck is also found (Grosberg and Strathmann, 1998, 2007; Bourke, 2011; Fisher et al., 2013). Complex multi-cellularity is not found in cases where multi-cellularity is achieved by cells coming together (aggregating) but only when multi-cellularity is achieved by a single initial cell dividing and staying stuck together with its offspring. (See Fisher et al.'s Figure 2 for a very useful colour-coded phylogenetic tree showing the clear correlation between clonal group formation through a bottleneck with obligate and complex multi-cellularity).

A unicellular bottleneck has arisen once in the animals (Dawkins, 1982; King, 2004), once in the land plants (Ligrone et al., 2012), three times in the fungi (the ascomycetes, basidiomycetes and chytridiomycetes; Niklas and Newman, 2013), once in the red algae (Coelho et al., 2007) and at least once (probably more) in the brown algae (Clayton, 1988; Charrier et al., 2008). For useful reviews, see Grosberg and Strathmann (2007), Bourke (2011, p.13), Knoll (2011), Fisher et al. (2013), Niklas (2014).

While the various evolutions of multi-cellularity from single-celled eukaryotes receives a lot of attention, there are analogous fraternal transitions at other levels in which bottlenecks are also implicated. A few examples are presented in what follows.

4.3.2 Eusocial insects employ a colony-level life-history bottleneck

The distinction between complex and simple multi-cellularity has an analogue in the social insects (Bourke, 2011). Some species are eusocial, with reproductive division of labour and caste dimorphism. In such species, colonies are usually founded by one singly-mated queen (Boomsma, 2009). In those species where this is not the case, there is evidence that it has been the case in that species' evolutionary past but has been lost secondarily (Boomsma, 2009).

Social insect colonies without a colony-level life-history bottleneck also lack reproductive division of labour and caste dimorphism, being analogous to the case of simple multicellularity (Bourke, 2011).

4.3.3 Naked mole-rats employ a colony-level life-history bottleneck

Although it is not yet explicitly recognised as such in the literature, I will claim in this section that the colony life-history of the naked mole-rat (*Heterocephalus glaber*) involves a clear example of a life-history bottleneck at colony level, analogous to that in the eusocial hymenoptera.

The naked mole-rat is a eusocial mammal[1] native to the Horn of Africa (Jarvis, 1981). Naked mole-rats live in subterranean colonies consisting of an average of 80 and up to 300 individuals (Jarvis, 1991). Theory about the origin and maintenance of eusociality in the naked mole-rat has undergone a marked shift in the past 20 years. I want to draw a distinction between early work in the eighties and nineties (where in-breeding as a mating system was used to do the explanatory work) with later work that is slowly coming to recognise that colonies may have a degree of Darwinian individuality in their own right. As will shortly become clear, bottlenecks of one sort or another have a role to play in both bodies of theory.

Earlier work: the in-breeding hypothesis

It has long been known that naked mole-rat colonies feature a reproductive division of labour, with only a single reproductive female (Jarvis, 1981). Generations overlap. Non-reproductives assist with parental care of the young, foraging for food and defense of the colony. Studies of genetic diversity in wild populations of naked mole-rats found that generic variation within colonies and between nearby colonies was extremely low, with almost all individuals tested being homozygous at loci known to be diverse in similar species of rodents (Reeve et al., 1990; Faulkes et al., 1990; Honeycutt et al., 1991). Based on this evidence, it was though that eusociality was enabled by unusually high levels of intra-colony relatedness achieved through consanguineous mating within colonies (Reeve et al., 1990; Faulkes et al., 1990; Sherman et al., 1991). It was thought that naked mole-rat colonies reproduced by fission, with fragments of existing colonies leaving to form new ones nearby (Brett, 1991). Given the fact of very high

[1]The only other known eusocial mammal is the related Damaraland mole-rat (*Fukomys damarensis*), with a geographic range limited to southern Africa (Sherman et al., 1991). It is thought that eusociality has evolved independently in these two species (Jarvis and Bennett, 1993; Faulkes et al., 1997).

within-colony relatedness and the greatly elevated mortality risk faced by an individual choosing to leave the colony in search of matings, it was reasoned that the inclusive fitness interests of individuals were best served by remaining in the natal colony and assisting with brood care of their very close kin (Sherman et al., 1992). This genetic explanation of eusociality in naked mole-rats was widely accepted for many years.

More recent work: in-breeding avoidance and the 'disperser caste'

More recently, the in-breeding hypothesis has come to be challenged in ways directly relevant to the topic of this chapter. The reproductive division of labour observed by Jarvis (1981) is now known to include both males and females, with well-supported physical caste dimorphism (O'Riain et al., 2000) observed in females. Female workers remain much smaller than the queen and morphologically pre-pubescent. Male workers exhibit low sperm count and motility (Bennett and Faulkes, 2000). Colonies typically contain a single reproductive female, one to three reproductive males and large numbers of their (largely non-reproductive) male and female direct offspring. There is also some evidence of social differentiation of workers into different castes, specialising in the performance of different roles (Lacey and Sherman, 1991; Bennett and Faulkes, 2000).

Of particular interest here, there is now good evidence for the existence of a rare disperser caste (O'Riain et al., 1996), specialised for leaving the colony, finding mates and establishing a new colony. In laboratory conditions, dispersers preferentially breed with non-kin captured from different colonies and avoid breedings with kin, even if they have never met (Ciszek, 2000). Dispersers have much larger amounts of body fat than non-dispersers, high levels of luteinizing hormone (a promoter of ovulation in females and testosterone production in males) and, in laboratory conditions and unlike other colony members, they continually try to escape from the box in which their colony has been housed (O'Riain et al., 1996). The disperser caste has been observed in wild populations and there is now evidence that establishment of new colonies by pairs of non-kin dispersers from different parent colonies is common (Braude, 2000). Recent population-genetic analysis (Ingram et al., 2015) based on data from a much a wider geographical area than early studies has shown that within-colony relatedness is no higher than what would be expected for half-sibs ($r \approx 0.25$), consistent with a single reproductive female mating with a small number of reproductive males.

I suggest that naked mole-rat colonies exhibit a life-history bottleneck at colony level. A disperser morph of the type discussed by O'Riain et al. (1996), Braude (2000) and Ingram et al. (2015) would instantiate a unitary propagule from the perspective of the mitochondrial genome and a small propagule from the perspective of the nuclear genome. Such a bottleneck is not as extreme as the case of multicellularity but nevertheless would have the effect of partitioning genetic variation between more than within colonies, leading to a partial shift in the level of Darwinian individuality. Character-fitness covariance: colonies emitting larger numbers of successful dispersers would be fitter than colonies emitting less. Heritability at colony level: colonies founded by a single female would presumably give rise to new colonies having the same mitochondrial genome (high heritability at colony level) and sharing half of the nuclear

genome with the parent colony. The fact of polyandry means that within-colony selection could occur (fathers may be unequally represented in the dispersers). Aside from a passing reference to multi-level selection in the context of 'further work' by Ingram et al. (2015), I am not aware of any interpretation of the population structure of naked mole-rats that interprets their colony-formation system as a life-history bottleneck (or a germ-line segregation).

Whence the unusually high homozygosity data found in the 1990's?
The genetic homogeneity observed in earlier work (e.g. Reeve et al., 1990; Faulkes et al., 1990; Honeycutt et al., 1991) was later found not to be due to in-breeding. It was due to the use of samples all taken from the same small region of Kenya - a region in which the naked mole-rats were all descendants of a small number of colonisers at the very edge of the species' geographic range (Braude, 2000; Ingram et al., 2015). This is a phenomenon known in population genetics as a *population bottleneck*[2]. This scientific wrong-turn is of interest because, despite operating in very different contexts and at different hierarchical levels, there is an important similarity between a population bottleneck and a life-history bottleneck - both reduce genetic variation in analogous ways (see boxed text on 'Bottlenecks And Sample Variance' in Section 4.4 below). So, while the proponents of the in-breeding hypothesis were wrong about the source of unusually high homozygosity, they were wrong in an interesting way.

4.3.4 Siphonophores employ a colony-level life-history bottleneck

Siphonophores are a phylum of colonial cnidarians (class hydrozoa) with problematic status as biological individuals (Beklemishev, 1969, pages 483-490). They are elongated marine animals of high morphological complexity, with a body plan stretched out along a line (Dunn, 2005). One of the 175 known species lives on the surface (the Portuguese Man O'War, *Physalia physalis*), a few live on the bottom (the *Rhodaliids*) but most occupy the pelagic zone, where they predate on small zooplankton (Dunn, 2009). They are difficult to study because they tend to disintegrate in nets or when experiencing the drop in pressure associated with traps being brought up to the surface from the deep (Mackie et al., 1987). They range in size from a few centimetres to tens of metres - some siphonophores are the longest known animals on Earth (Dunn and Wagner, 2006).

Siphonophores have the appearance and behaviour of complex organisms but are in fact colonies of clonally-related zooids, each zooid being an animal homologous with other solitary-living

[2]In population genetics, the term 'population bottleneck' is used to refer to an event in which a population has been founded by a relatively small number of individuals (Nei et al., 1975). This 'founder event' can be due to a migration event, a near-extinction event or a geographical event that reproductively isolates a small fragment of a larger population. It is of significance in population biology (and anthropology) because such population bottlenecks cause an increase in homozygosity that persists for many generations. If the population bottlenecking event was in the distant past and the size of the focal population has since grown larger, the homozygosity detectable in the population will be higher than that expected from the population size and mutation rate, so providing a detectable signal of past events in the allele frequencies of the present.

cnidarians. Zooids are physically attached to one another along the main axis and share a common vascular system, digestive system and coordinated goal-directed behaviour. Colonies behave as a functionally integrated whole, with striking division of labour and specialisation of parts. Specific functional roles are associated with each point along the main axis, with zooids specialising in somatic functions such as buoyancy control (pneumatophore), swimming (nectophores), catching prey, digesting prey (gastrozooids) and others (Mackie et al., 1987). There is also reproductive division of labour among zooids, to which I return shortly.

The ontogeny of siphonophore colonies is of interest here, as I suggest it employs a life-history bottleneck at colony level. (I am not aware of any connection made in the literature between what is known about the development of siphonophore colonies and the life-history bottleneck phenomenon.) The colony life cycle includes a diploid, sexually produced zygote: a single-celled fertilised egg from which all zooids in the adult colony will be descended. This develops into a larval stage that is a single zooid known as a protozooid. This buds off a clone and the connection between the two elongates into a stalk along which further clones will bud off in between them according to a linear body plan. Zooids epigenetically differentiate into different functional types according to a spatial program, so that zooid development is canalised according to sequential position along the linear axis of the developing colony (Mackie, 1986).

The division of labour includes reproductive division of labour. A small minority of zooids differentiate into the gonozooids, the parts of the colony involved in colony-level reproduction. There is thus germline segregation in addition to a life-history bottleneck[3]. In this sense, the non-reproductive zooids display so-called reproductive altruism.

One might object that there is a sense in which the single-zooid bottleneck simply piggy-backs on the uni-cellular bottleneck instantiated by the zygote from which the proto-zoid develops. The zooids are, after all, complex eukaryotic multi-cellular organisms in their own right and are homologous with other non-colonial hydrozoans that develop from single-celled bottlenecks. While this is true, it must be remembered that this re-use of pre-existing machinery is common to all hierarchical transitions. For example, it is also the case for eusocial hymenopteran colonies, where there is wide acceptance of 'organismality' status at colony level (Wilson and Sober, 1989; Queller and Strassmann, 2009), that colonies are founded by propagules (mated queens) that are themselves the products of an ontogenetic process very similar to that in solitary

[3]Bourke's (2011) 'virtual dominant' hypothesis for the evolution of an early-segregating, slowly-dividing germline in multi-cellular animals might be tested by applying it to colony development in siphonophores. According to Bourke, the average relatedness of somatic cells to germ cells is greater than that between somatic cells (due to mutation distances in the tree-like cell lineage that is the animal) and this explains why it is in the inclusive-fitness interests of somatic cells to further the germline cells' reproduction rather than their own. If Bourke's theory applies in general then one would expect to find that the gonozooids of siphonophores are at a lower budding-distance from the proto-zooid than the non-reproductive somatic zooids. Casey Dunn (personal communication) reports that the gastrozooids (not the gonozooids) are the first to bud off the proto-zooid. Since budding occurs along a linear axis, that gives the gastrozooids the least mutation distance from the proto-zooid. He also reports that insofar as there is a pattern, in larger species the gonozooids differentiate later in ontogeny (Bourke's virtual-dominant considerations suggest that early differentiation becomes more important with larger size.) Furthermore, Dunn reports that it is unknown whether the cells in the gonozooids divide more slowly than other zooid types. (Bourke's virtual-dominant considerations suggest that slow division of germ cells becomes more important with larger size.) This would be an interesting topic for empirical study.

hymenoptera. For a second example, mitosis is a similar process in unicellular and multicellular eukayotes. I argue that it is normal that major evolutionary transitions re-use the reproductive machinery of lower levels, albeit organised in a new way and that such re-use of machinery is not evidence against a change in the level of Darwinian individuality.

4.3.5 Metazoan cells employ a mitochondrial DNA bottleneck

Mitochondrial DNA is inherited uni-parentally in mammals (and most other phyla), being passed down the female line (Wolstenholme, 1992; Birky, 1995). Its molecular characteristics and location in the cellular cytoplasm mean mitochondrial DNA is subject to significant mutation rates (Avise, 1991). As a dividing haploid and asexual lineage, mitochondrial DNA is prone to the accumulation of deleterious mutations known as Muller's Ratchet (Muller, 1932; Charlesworth et al., 1993). Mitochondrial DNA mutations are known to be common (Lynch, 1996) with cells bearing genetic diversity in their mitochondrial DNA said to suffer from heteroplasmy. The deleterious effects on cell fitness increase monotonically with the degree of heteroplasmy. That mitochondria have lived in this way since (at least) the evolution of the eukaryotic cell without suffering mutation meltdown therefore stands in need of explanation.

What is known is that the copy number of mtDNA in an oocyte (egg) is of the order of 10^5. The oocyte may bear heteroplasmy. After syngamy and just a few rounds of cell division, it is known that the daughter cells bear a heterogenous distribution of heteroplasmy with high variance. (That is, some cells bear a lot of heteroplasmy, some little or none). This exposes mutant mtDNA to between-cell selection, acting to reduce its frequency in the cellular population by removing cells containing higher heteoplasmy. There exists substantial controversy about how this is achieved. (For a review, see Johnston et al., 2015).

One of the most prominent hypotheses is the 'mtDNA bottleneck hypothesis', due to Bergstrom and Pritchard (1998). They postulate a mechanism as follows. At the beginning of ontogeny, just after syngamy, the oocyte (now a zygote) contains around 10^5 instances of the mtDNA, of which a certain proportion will be mutants. Once cell division starts, the number of cells multiplies but the mtDNA does not replicate. Its copy number remains constant for the first few rounds of division (of the order of 10 rounds), such that it is spread ever more thinly between the daughter cells. mtDNA are allocated to daughter cells randomly following division. After a certain small number of divisions, the mtDNA starts replicating and its numbers rise again until they return to and stabilise at around 10^5 per cell.

Clearly the extent of mtDNA bottlenecking will increase with the number of rounds of cell division for which mtDNA replication is suppressed in such a model. Likewise, the between-cell variance in heteroplasmy will increase with the extent of bottlenecking. If between-cell selection against heteroplasmy obtains then it will act in the role of purifying selection, removing deleterious mutations from the mtDNA lineage.

While this simple model has been superseded by many others more faithful to the messy details of the cytology, its essence has been retained by several (see references in Johnston et al., 2015) and it seems likely that some sort of bottlenecking-related sample variance phenomenon is at work.

I should acknowledge that, while the bottlenecking effect (increased variance in small samples) is similar to the other examples, it seems likely that the mtDNA bottleneck is applied exogenously by the host cell and is therefore perhaps not a good example of the mtDNA being the 'subject and object of its own social evolution'.

4.3.6 The preceding examples show that a life-history bottleneck is a multiply realizable life-history trait

The message I take from these diverse examples of bottlenecks is that bottlenecking is a multiply realizable life-history trait for a collective. As life-histories go, it is quite simple to implement. It occurs in sexual and asexual contexts, eukaryotes and prokaryotes, multicellular organisms and colonies of them.

4.4 A life-history bottleneck is an assortment-driven Social Niche Modifier

There are many hypotheses for the evolution of life-history bottlenecks. Multi-cellularity presumably evolved because larger cell number was adaptive. Plausible sources of adaptive advantage include economies of scale, avoidance of predation and access to new niches dependent on large size (e.g. Lurling and Beekman, 2006; Velicer and Vos, 2009). The question is why, given the advantage of large size at one point in the lifecycle, it should also be adaptive to employ such a minimal size in another part of the lifecycle?

Perhaps the most prominent involves the clonal relatedness it conveys on cells (or high relatedness on colony members), so acting to discourage within organism/colony conflict (Queller, 2000). An assortment-driven view of this sort is the one I will pursue in this thesis. I pause here to note that there are numerous other hypotheses for the evolution of life-history bottlenecks, most of them mutually compatible. Some have suggested that bottlenecks exist because they facilitate cross-over as part of sexual reproduction (Bonner, 2000). Others have suggested they exist because they expose deleterious mutants to purifying selection. This effect is particularly important in asexual lineages that would otherwise be susceptible to the accumulation of deleterious mutations known as Muller's Ratchet (Klekowski, 1988; Grosberg and Strathmann, 1998). Another hypothesis concerns dispersal - generally smaller propagules are easier to disperse (Bonner, 2000). Rick Grosberg (personal communication[4]) has suggested a life-history

[4]Cooperation and Major Evolutionary Transitions conference, UC Santa Barbara, February 2013

bottleneck can act as an 'epigenetic reset button', allowing for a cyclical developmental programme that starts again with each new generation. Yet more hypotheses have been put forward (e.g. Rainey and Kerr, 2010). See Grosberg and Strathmann (1998, 2007) for a review of the candidates.

In this section I focus on the idea that collective life-forms face social dilemmas of the type discussed in Chapter 3 on Social Niche Construction. **I investigate how a bottlenecked life-history modifies the social niche experienced by particles living in that population structure by increasing the assortment of social behaviour.**

Bottlenecking can be understood as a Social Niche Modifier that alters the social niche experienced by particles by increasing the probability of like strategies meeting like. The limits are easy to understand intuitively. If an imaginary propagule were the same size as its parent collective[5] or comparable in size to its parent - then the strategy variance in the propagule would be equal to that in the parent. (Hence the probability of like strategies meeting in the offspring would be the same as they were in the parent, any within-collective variation would be heritable and multi-generational between-particle character-fitness covariance would be possible - so the social niche would be unmodified.) At the opposite extreme, if the propagule were to be a single particle then clearly the strategy variance within the newly founded collective would be zero (so like strategies would always meet). We can ask how bottlenecking affects assortment between these two extremes.

In the bottleneck model, the probability of like strategies meeting (call this $Pr(X, X)$) in randomly drawn pairs of particles within a propagule is as follows (see boxed text for derivation):

$$Pr(X, X) = 1 - \left(\frac{n-1}{n}\right)\left(1 - (p^2 + (1-p^2))\right) \tag{4.4.1}$$

where p is the proportion of cooperators in the parent collective and n is the bottleneck size (the number of particles the parent contributes to the propagule). See Figure 4.4 for a plot. (Note that in the model specification in Section 4.5 below, I use notation n^0 for bottleneck size, to distinguish it from the number of particles in a collective at maturity, n^t. There is no need to make that distinction here, so I use simply n.)

Bottlenecks and Sample Variance

Bottlenecking elevates the probability of like strategies meeting as a result of the phenomenon, well known in statistics (Brown, 1947), that the variance in a sample taken from a statistical population is systematically less than the variance of the population from which it was taken. For that reason, statistical sample variance is said to be a 'biased estimator' of population variance, with sample variance a fraction $(n-1)/n$ of the variance in the population (where n is the sample size). In the

[5]This thought experiment perhaps stretches the notion of what a propagule is - this would be more like a temporal continuation of the parent.

present case, the particles in a propagule are effectively a sample drawn from their parent collective and the variation in the propagule will be less than that in the parent collective by a factor of $(n-1)/n$.

Working in the population genetics tradition, Nei and Roychoudhury (1974) applied the small-sample effect to estimate the expected homozygosity at a single diploid locus in a small sample of size n taken from a random-mating population containing two alleles at the locus (having frequencies p and $1-p$). Their reasoning can be understood as follows. The expected population homozygosity, j, is

$$j = p^2 + (1-p)^2$$

The population heterozygosity, H, is then

$$H = 1 - j = 1 - \left(p^2 + (1-p)^2\right)$$

The expected sample heterozygosity h will be a fraction of the population heterozygosity (due to the small-sample effect)

$$h = \left(\frac{n-1}{n}\right) H$$

The expected sample homozygosity, g, is then $1-h$, i.e.

$$g = 1 - \left(\frac{n-1}{n}\right)(1 - (p^2 + (1-p)^2))$$

The same expression may be used to get the probability of two randomly drawn particles having the same strategy (call this $Pr(X,X)$) when those two particles are drawn from a propagule that is itself a sample of size n drawn from a parent collective containing two alleles having frequencies p and $1-p$. In the present case, p is the frequency of cooperators in the parent, $1-p$ is the frequency of defectors in the parent and n is the absolute bottleneck size, measured in particles.

Thanks to Guy Jacobs (personal communication) for pointing me to Nei and Roychoudhury's result. See also Hedrick (2011, p.100).

In Chapter 3 we met a measure of assortment, α, defined such that each particle plays its own strategy with probability α and with probability $(1-\alpha)$ it plays against the strategy of a randomly chosen member of the collective, whatever that might be. This is a useful measure (or driver, depending on context) of assortment because it allows us to calculate the effective game being played on the T-S plane by a population facing a base game (T,S) in the presence of assortment α (see Section 3.8.3). Importantly, α is an abstraction away from whatever circumstances actually caused the assortment, allowing us to give general treatment of what assortment does to social niches regardless of how it arises. In this chapter, it will therefore be useful to understand how the imposition of a bottleneck of a certain size n changes assortment α. We can do this by writing down the probability of like strategies meeting in randomly drawn pairs of particles as a function of alpha and equating this with the same probability as a function of bottleneck size n (Equation 4.4.1). Recall that the ambient proportion of cooperators is p and of defectors is $1-p$.

The probability of a pairwise interaction resulting in mutual cooperation or mutual defection is, in terms of assortment α:

$$Pr(C,C) = p(\alpha + (1-\alpha)p)$$

$$Pr(D,D) = (1-p)(\alpha + (1-\alpha)(1-p))$$

The probability of like strategies meeting is then the sum of these two terms, which simplifies to:

$$Pr(X,X) = -2\alpha p^2 + 2\alpha p + 2p^2 - 2p + 1 \tag{4.4.2}$$

This gives us the expected frequency of like strategies meeting, given a certain proportion of cooperators, p, and a certain imposed level of assortment, α (see Figure 4.5 for a plot). This expression can be combined with Equation 4.4.1 to find α as a function of n:

$$\begin{aligned} Pr(X,X) &= 1 - \left(\frac{n-1}{n}\right)\left(1 - (p^2 + (1-p)^2)\right) \\ &= -2\alpha p^2 + 2\alpha p + 2p^2 - 2p + 1 \end{aligned}$$

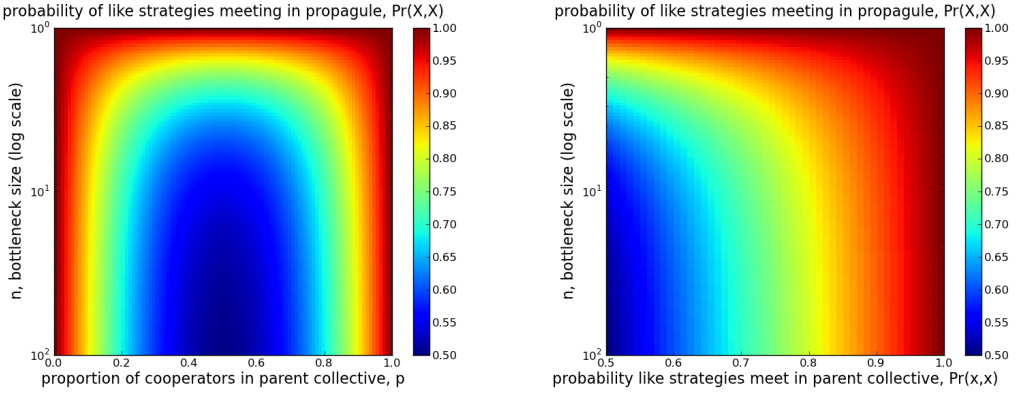

FIGURE 4.4: **Effect of bottlenecking on $Pr(X,X)$:** In the bottleneck model, the expected frequency with which like strategies are paired in an offspring propagule, $Pr(X,X)$, depends on the frequency of cooperators in the parent collective, p and the bottleneck size used in forming the propagule, n (LEFT), as per Equation 4.4.1. The probability of like strategies meeting in the parent collective is $Pr(x,x) = p^2 + (1-p)^2$. The bottlenecking process can be seen as increasing the probability that like strategies meet (RIGHT). Note 1 - Bottlenecking makes the most difference when there is maximum strategy variability ($p = 0.5$) and makes no difference when there is no strategy variability ($p = 0$ or $p = 1$) in the parent. Note 2 - The homogeneity caused by bottlenecking is a non-linear function ($1/n$) of bottleneck size, having little effect at low bottleneck severities (unlike α which is a linear difference-maker by definition).

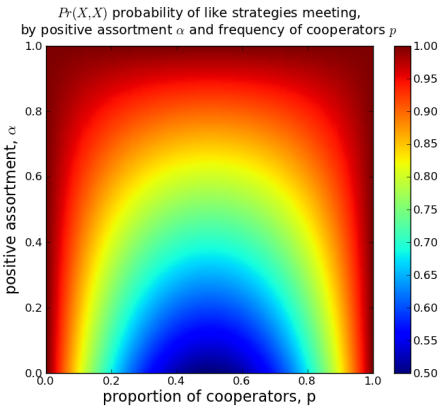

FIGURE 4.5: **Effect of imposed assortment α on $Pr(X, X)$:** In generalised treatments, the expected frequency that like strategies meet, $Pr(X, X)$ depends on the p, the background proportion of cooperators and α, the imposed level of assortment. (Equation 4.4.2). Note 1 - Imposed assortment level, α, makes the most difference when there is maximum strategy variability ($p = 0.5$) and makes no difference when there is no strategy variability ($p = 0$ or $p = 1$). Note 2 - The homogeneity caused by imposing assortment is by definition linear in α.

Simplifying, we find the assortment α as a function of bottleneck size n is simply

$$\alpha = \frac{1}{n} \tag{4.4.3}$$

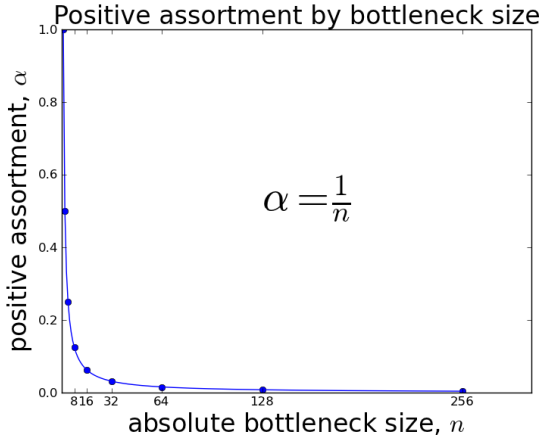

FIGURE 4.6: Assortment, α, is inversely proportional to absolute bottleneck size, n (Equation 4.4.3). Bottlenecking has a large effect on assortment only when a small number of particles pass through the bottleneck. The relationship does not depend on the size of the parent collective.

Figure 4.7 illustrates this effect clearly, by plotting points on the T-S plane. It shows how the effective game (the modified social niche) for particles in offspring propagules differs from the base game (social niche) that obtained in their parent collective, according to the bottleneck size used in creating those propagules.

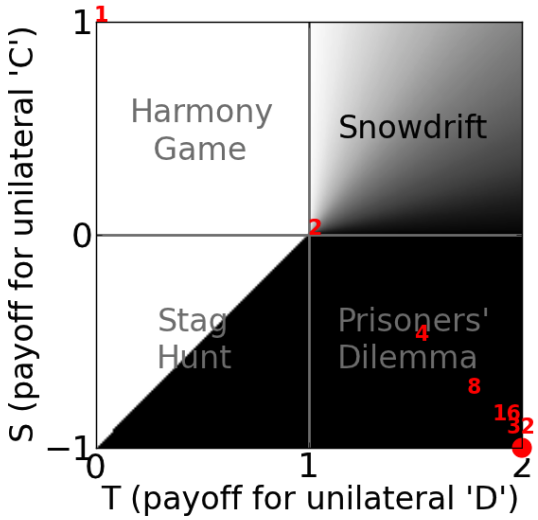

FIGURE 4.7: **Bottlenecking is a source of assortment** (Equation 4.4.3). A single round of bottlenecking alters the effective game being played (the social niche) relative to the game that would have obtained without bottlenecking. The position of the red numbers on the T-S plane show the effective game for that value of bottleneck size (n), given the base game shown as the red dot (T=2,S=-1; a Prisoners' Dilemma). For legibility, only bottlenecks $n \in \{1, 2, 4, 8, 16, 32\}$ are shown. Larger bottlenecks have a decreasing effect on the effective game. A unitary bottleneck guarantees a Harmony Game for any base game. Importantly, the analysis here is for a single round of bottlenecking (one generation). Repeated applications of a bottleneck have a cumulative effect.

There are a few things to note. Firstly, while α by design provides a linear scale from no assortment to full positive assortment, bottlenecking is clearly a non-linear cause of positive assortment (Equation 4.4.3). This means the bottleneck really only has a significant effect when the bottleneck is severe - that is, when the number of particles in the propagule is small. Secondly, like any assortment modifier, a bottleneck makes the most difference when there is abundant available variation in the parent for it to act upon and makes no difference when there is no variation (see Figures 4.4 and 4.5). So assortment-causers like bottlenecks are difference-makers when $p = 0.5$ but have very little effect when one strategy is rare. (This is important for understanding the case of rare mutants.) Thirdly, the sampling effect is not dependent on the size of the 'population' (in this context that means the size of the parent collective) from which the sample is drawn. This means that absolute and not relative bottleneck size is the cause of elevated strategy assortment (and hence social niche modification) in the bottlenecking case. Fourthly, the analysis here is for a *single round* of bottlenecking. Repeated applications of a bottleneck will have a cumulative effect - this will be important when interpreting the computational model presented below in which bottlenecking is intercalated with rounds of within-collective selection, mutation and drift.

4.5 A model of the evolution of a life-history bottleneck by social niche construction

This section presents a simulation model for the evolution of a life-history bottleneck. It demonstrates an evolutionary transition in individuality (as defined in Section 1.4) by social niche construction. It assumes a two-level population structure, with particles nested within collectives (Figure 4.8). It is not intended to explain how such a group structure came about. Particles are

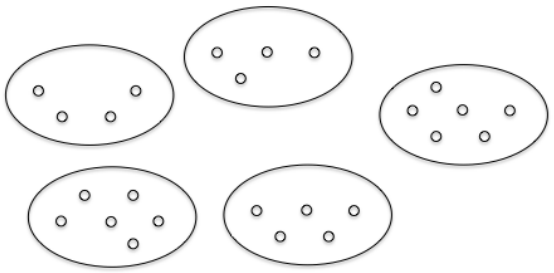

FIGURE 4.8: The Bottleneck Model employs a two-level structure, with particles nested inside collectives.

asexual, haploid and have two heritable traits.

- Social Trait - 'Cooperate' or 'Defect'
- Social Niche Modifying Trait - Bottleneck Size Preference

Firstly, particles have a Social Trait controlling their resource consumption strategy - either 'Cooperate' or 'Defect'[6]. Secondly, they have a Social Niche Modifying Trait which influences the life-history of the collective of which the bearer is a part. The collectives employ propagule emission as a reproductive strategy - they allow parts of themselves to break off and disperse (as in Wallace's comments in the boxed text in Section 4.2 above) before dying. The Social Niche Modifying Trait in this model is called Bottleneck Size Preference because it affects the size of the fragments that break off the parent collective to found new ones. The Bottleneck Size Preference locus has eleven 'alleles', each mapping onto a particular preferred number of particles in the bottleneck, ranging from a single particle to a large number of particles (on a log scale, enumerated in Table 4.1). It is a 'preference' because a single particle usually does not have the power to determine the propagule size of the collective of which it is a part. The size of the propagule can be influenced by all the particles within each collective and they may vary in their preference. Particles are otherwise undifferentiated.

[6]The safety quotes are a reminder of the need to consider the terms 'Cooperate' and 'Defect' as having their common connotation only sometimes - as discussed in Section 3.8.3.

4.5.1 Life cycle overview

The Bottleneck Model employs a Wright-Fisher style model design (see Chapter 6 for a discussion of this class of evolutionary models). Generations do not overlap. The total number of collectives is held constant and collectives compete for representation in the subsequent generation. The model is initialised with a population of N collectives, each initially having the same number of particles and the same distribution of particle types.

Model execution involves cyclically iterating an algorithm. Each iteration represents one generation, counted by T. In each generation, there is first a collective Growth Phase in which particles multiply by selfing within their respective collectives, at a rate proportional to their payoff in a game representing their within-collective social niche. There is then a collective Reproduction Phase in which all collectives simultaneously mature, emit a number of offspring collectives and die. The expected number of offspring collectives emitted by each collective is a linear function of how many particles are in it. Fatter collectives are fitter collectives. The total number of collectives is held constant at N and collective generations do not overlap. Both the Social Trait and Social Niche Modifying Trait (Bottleneck Size Preference) are subject to mutation at low rates, with the Social Trait changing more rapidly than the Social Niche Modifying Trait. The offspring collectives form the next generation and the cycle repeats for a large number of generations. Due to mutation, selection and drift, collectives can differ in their proportion of cooperator particles and in their life history bottleneck.

Because bottlenecking is applied to successive generations, intercalated with episodes of intra-collective particle-level selection, mutatation and drift, it is not possible (for me) to derive an analytic prediction of how the model will behave. The intention of the model is to provide an experimental system to test intuitions, to serve as an existence-proof for a system in which social niche construction of a life-history bottleneck occurs and to examine the effect of bottlenecking on the hierarchical level at which character-fitness covariance obtains (using the Price Approach (Clarke, 2010)).

4.5.2 Growth Phase - collectives grow by vegetative multiplication of their particles (subject to a collective action problem)

The collective action problem facing particles within each collective is modelled using a pairwise two-player, two-strategy symmetric game. I call this the 'base game' and this is the unmodified social niche experienced by particles. (We will see shortly that the 'effective game' played within each collective may differ from the base game as a result of the aggregate effect of the Social Niche Modifier traits of the particles concerned - and that this differing from the base game may vary across collectives.) The per-capita rate of increase of each type (Cooperate and Defect) of particle is determined by the payoff to each strategy in this game[7]. The game can be any game on the T-S plane (Figure 3.1, page 44) and is supplied to the model as a parameter, in

[7]The Social Niche Modifier trait has no direct effect on a particle's per-capita rate of increase.

the form of a payoff matrix (R, S, T, P). Recall from Table 3.2 (page 43) that R is the payoff for mutual cooperation, S for unilateral cooperation, T the payoff for unilateral defection and P the payoff for mutual defection. The base game is constant across all collectives during a simulation run. It is interpreted as representing the facts of (social) life for a particular biological situation, as discussed in Section 3.8.1. A constant β, greater than the maximum possible payoff, is added to all payoffs - this implements weak selection[8].

Let p be the proportion of cooperator particles in a focal collective. The expected payoffs to individual cooperator and defector particles in that collective are then, respectively:

$$E(\pi_C) = pR_+(1-p)S + \beta$$
$$E(\pi_D) = pT_+(1-p)P + \beta$$

(The particles' Social Niche Modifier trait - their Bottleneck Preference - does not directly affect their expected payoff.) Particles reproduce within their collectives, with game payoff used directly as the per-capita rate of increase of each type:

$$w_C = E(\pi_C)$$
$$w_D = E(\pi_D)$$

Gameplay and clonal particle reproduction are repeated for t iterations, where t is usually 1. (This is because modelling within-collective dynamics with more than one round of gameplay ($t > 1$) would introduce a second source of assortment, because like particles tend to beget like and particles do not 'disperse' within the lifetime of their containing collective (Fletcher and Zwick, 2004). As this model is used to investigate the social niche modifying effects of a bottlenecked life history, this additional source of assortment would introduce a confounding variable and so is deliberately excluded.)

Particles usually breed true, meaning they usually beget particles similar to themselves in both their Social Trait and Social Niche Modifier (Bottleneck Size Preference) trait. However, in each case it is possible that a mutation may occur (with probabilities μ_{ST} and μ_s respectively), introducing a small amount of random variation.

The number of particles in collective k at inception is denoted n_k^0 and the proportion of cooperators at inception p_k^0. After the Growth Phase has completed, collectives are said to have reached maturity. The number of particles in collective k is then denoted n_k^t and the proportion of cooperators p_k^t. Note that the mechanism described in this section implies that particles have heritable traits that can vary and that affect their relative reproductive success in within-collective dynamics.

The algorithm then moves to the Reproduction Phase.

[8] Adding a constant to all entries in a payoff matrix does not affect the equilibria of the game, only the strength of selection moving toward equilibrium (Weibull, 1995; Hammond, 2005; Colman, 2013, p.78).

Notation used in Bottleneck Model

- N total number of Collectives in population
- k index for collectives
- index k' indicates a collective that is an offspring of collective k
- i index for particles within global population of particles
- j index for particles within a single focal collective
- t number of rounds of gameplay in each generation (usually 1)
- (R, S, T, P) the payoff matrix of the 'base game' determining the *unmodified* social niche experienced by all particles
- T generation counter (time)
- Particle properties:
  - ST Social Trait locus, $ST \in \{$'Cooperate', 'Defect'$\}$
  - s Bottleneck Size Preference locus, $\{s \in \mathbb{Z} | 0 \leq s \leq 10\}$. The Social Niche Modifying trait in this model, affecting bottleneck size of the containing collective, S_k.
  - μ_{ST}, μ_s Mutation rates for Social Trait and Bottleneck size respectively
- Collective properties:
  - n_k^0 number of particles in collective k at start of generation (i.e. **bottleneck size**)
  - p_k^0 proportion of Cooperators in collective k at start of generation
  - n_k^t number of particles in collective k at end of generation (at maturity)
  - p_k^t proportion of Cooperators in collective k at end of generation (at maturity)
  - w_C, w_D fitness (per-capita rate of increase) of particles bearing Social Trait 'Cooperate','Defect' respectively. This is considered on a per-collective basis, so strictly w_{k_C},w_{k_D} but usually context allows omission of the k subscript. Depends on payoff. Measured over a single generation.
  - S_k Bottleneck Size employed by collective k where $\{S \in \mathbb{Z} | 0 \leq S \leq 10\}$. Collective k emits propagules of size $n_{k'}^0 = 2^{(S_k)}$, so the expression of $S_k = 10$ causes propagule size $n_{k'}^0 = 1024$ while the expression of $S_k = 0$ causes $n_{k'}^0 = 1$ (i.e. a unitary propagule). See table 4.1.
  - Z_k mean cooperativeness of particles in collective k at inception (Cooperators have cooperativeness 1 and defectors 0, so $Z_k = p_k^0$)

– W_k the number of offspring collectives or 'propagules' begot by collective k.

(In subsequent discussions, collective fitness understood as number of offspring collectives (like W_k here) will be called collective fitness$_2$ while collective fitness understood in terms of offspring particles will be called collective fitness$_1$.)

4.5.3 Reproduction part 1: establish collective-level fitness

Each generation, after the Growth Phase the collectives enter the Reproduction Phase. First the fitness$_2$ (number of offspring collectives) of each collective is determined, then offspring collectives ('propagules') are created and they become the next generation.

The total number of collectives is held constant. Slots in the subsequent generation are allocated to collectives with a probability determined by comparing their size (their particle number at maturity, n_k^t, abbreviated here to n_k for legibility) with the mean collective size, $\bar{n}$. The expected fitness$_2$ of focal collective k is then $E(W_k) = n_k/\bar{n}$. The model uses fitness proportionate selection to convert this expected fitness (a real number) into an actual number of offspring, W_k (an integer). Collectives with above-average particle number are more likely to have multiple offspring collectives ($W_k > 1$); collectives with below-average particle number are more likely to have no offspring collectives ($W_k = 0$).

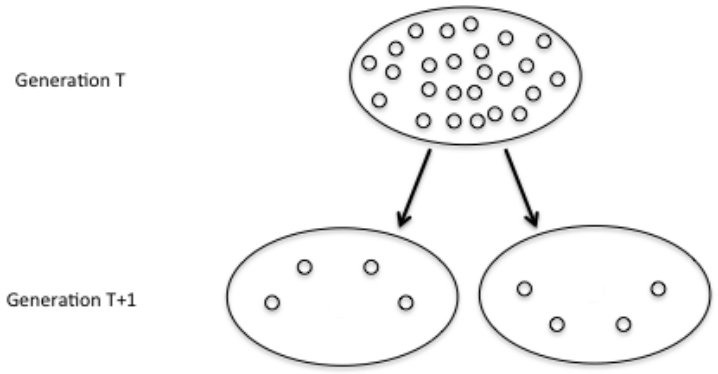

FIGURE 4.9: Here we see a focal collective (for which $W_k = 2$) beget two offspring collectives. The character of each particle (ST and SNM traits) is determined by sampling from the parent. In this example, the collective's Bottleneck size $S_k = 2$, so each propagule contains $n_{k'}^0 = 2^{(S_k)} = 4$ particles. The particles' Social Trait is not shown.

Note: (Damuth and Heisler, 1988) distinguish two types of model of multi-level selection. In MLS1 models, particles are grouped into collectives but only particles participate in the parent-offspring relation; collectives in MLS1 models are just a population structure imposed on the particles and are not tracked in their own right. In MLS2 models, both particles and collectives

participate in the parent-offspring relation: particles beget particles and collectives beget collectives. Now consider collective fitness in the two cases. In MLS1 models, collective fitness can only be measured in terms of the number of particles each collective contributes to the global particle population - call this fitness$_1$. In MLS2 models, collective fitness can still be measured in terms of the number of particles each collective contributes to the global particle population - also call this fitness$_1$. But MLS2 models also admit of another understanding of collective fitness, measured in terms of number of offspring *collectives* - call this fitness$_2$. The model presented here is an MLS2 model with collective fitness$_2$ proportional to collective fitness$_1$. (The significance of this will become clear in Chapters 5 and 6.) Notation follows Okasha (2006).

4.5.4 Reproduction part 2: create offspring collectives

Once the number of offspring collectives, W_k, for each collective has been determined, the content of those offspring is then established.

Determine propagule size

The size of the offspring collectives, if any, is affected by the Bottleneck Size Preference traits of all the particles in the parent collective. This is done by selecting one particle at random from the parent collective and using the value of that particle's Bottleneck Size Preference trait (s_j) to determine the actual extent of bottlenecking employed by the collective. This simple lottery method means the most frequent value of the Bottleneck Size Preference trait is most likely to be used. The Bottleneck Size Preference locus has eleven discrete 'alleles', integers in the range 0 to 10. If focal collective k has n_k^t particles at maturity and collective Bottleneck Size $S_k = s_j$ has been decided upon, then the absolute number of particles in a propagule emitted by k (call this offspring collective k') - is given by:

$$n_{k'}^0 = 2^{(S_k)} \tag{4.5.1}$$

The superscript ^{0} denotes the number of particles in the offspring collective k' at its inception (prior to its own subsequent growth). The genotype-phenotype mapping is arbitrary and could have been any 'lookup table'. An exponential mapping is used as a compact way to allow Bottleneck Size Preference to span a wide range with a manageable number of 'alleles' (Table 4.1) but S itself has no numerical significance (letters or any other symbols could also have been used).

Note also that the number of particles in the bottleneck is not relative to the size of the parent but is an absolute number of particles. This is because the effect of small samples on sample variance (discussed in Section 4.4) does not depend on the size of the statistical population from which the sample is drawn. After bottlenecking, all propagules are normalised (maintaining the distribution of particle types) to the same size (1024 particles). This is to remove the confounding variable of initial propagule (collective) size - I am interested in the effect of bottlenecking on sample variance (and not in the tradeoff between large numbers of small propagules and

small numbers of large ones). A similar normalisation is used in Bergstrom and Pritchard's (1998) model of mtDNA bottlenecks discussed in Section 4.3.5 above.

TABLE 4.1: Genotype-phenotype map: in the bottleneck model, the Social Niche Modifier is a Bottleneck Size Preference 'trait', expressed as a preference for propagules having a particular number of particles.

Bottleneck trait, S_k	**Particles in propagule,** $n^0_{k'}$
0	1 (unitary propagule)
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024

Select particles for inclusion in propagule

Having decided the absolute number of particles to include in the propagule, the next step is to select that number of particles from the parent. (Recall that there may be variation among particles, due to mutations occurring during the Growth Phase in this or previous generations.) Particles are stochastically selected (with replacement[9]) from the parent collective with a probability proportional to the similarity between the particles' Bottleneck Size Preference trait and S_k, the collective's bottleneck severity. This ensures a degree of assortment of the Social Niche Modifier. As discussed in Chapter 3, SNM-assortment is often an important factor in social niche construction, depending on the base game.

Importantly, the particles' Social Trait does not affect their chances of being selected for inclusion in the propagule and no assortment on Social Trait is coded explicitly into the model. Post-bottleneck normalisation means there is no intrinsic size-based advantage to any particular bottleneck size.

The offspring collectives form the next generation, the current collectives are discarded and the cycle repeats.

[9] One might reasonably object that a particle could only be included in a propagule only once, so selection without replacement might be more appropriate here. The use of selection with replacement simplifies implementation and is a good approximation for large n^t_k (for the same reasons that employing the binomial distribution as an approximation for the hypergeometric distribution is acceptable for large population size.) And n^t_k is indeed large (> 1024).

4.5.5 Measuring the level of selection using the Price Approach

The Price Approach (Section 2.3.3, adapted from Clarke (2010)) can be used to measure the level at which character-fitness covariance obtains for a single generation. A time-series record of it can be used to represent how it changes (diachronically) over the course of a simulation run. Recall that Clarke defines a continuum from 0 (representing wholly lower-level selection) to 1 (representing wholly higher-level selection). Recall also that I have modified Clarke's version of the Price Approach to take account of heredity, so that the Price Approach can be used not just to measure the level of selection but the level of evolution. As in Section 2.3.3 Equation 2.3.2, the relative strength of collective-level selection on the Price Continuum is then:

$$\frac{|cov(W, Z')|}{|cov(W, Z')| + |E(cov_k(w, z'))|}$$

When this is low, particle-level individuality is in force. When this is high, collective-level individuality is in force. When this changes from low to high, we have an evolutionary transition in individuality (as I have defined that term in this thesis).

Applied to the Bottleneck Model, this can be interpreted as follows:

- W is the vector (indexed by k) of the $E(W_k)$ values (one value per collective, N values). (This is both collective fitness$_1$ and expected collective fitness$_2$ in this model: the two are the same).
- Z' is the vector (indexed by k) of the mean cooperativeness values of the offspring collectives at the start of the subsequent generation (one value per collective in current generation, N values). The cooperativeness of collective k's offspring is the proportion of cooperators in the offspring collective at inception, $p^0_{k'}$. Where a collective has more than one offspring collective, the mean of its offspring's cooperativeness values are used.
- $cov(W, Z')$ then yields a single value
- $cov_k(w, z')$ has one value per collective (indexed by k)
- w is the vector of the fitnesses (per-capita rate of increase) of each of the particles in collective k, measured over a single generation. Each element will be one of two distinct values. One value is calculated for cooperators ($w_{kC} = \frac{n^t_k p^t_k}{n^0_k p^0_k}$) and the other for defectors ($w_{kD} = \frac{n^t_k (1-p^t_k)}{n^0_k (1-p^0_k)}$) (Section 4.5.2). These values are repeated in the vector w as many times as there are particles of that type at the start of a generation (total n^0_k values - defectors first, then cooperators).
- z' is the vector of the mean value of the Social Trait of the offspring of particles in collective k (again, total n^0_k values). One arbitrary value is used for cooperators (1) and another for defectors (0). These values are repeated in the vector z' as many times as there are particles of that type at the start of a generation (total n^0_k values - defectors first, then cooperators).

- $E(cov_k(w, z'))$ is then the mean of all of the $cov_k(w, z')$ values (yielding a single value).

Note that both particle and collective fitnesses are measured over the same time period - a single generation.

Using this method, a single value for the magnitude of covariance of fitness with both character and offspring character is logged, at both particle and collective level, for each passing generation. The results are presented below.

4.6 Simulation: The effect of bottleneck size on equilibrium frequency of cooperation, for a variety of base games

Before investigating social niche construction of a life-history bottleneck, I first test the effectiveness of an exogenously imposed life-history bottleneck as a Social Niche Modifier. The intuition is that the equilibrium level of cooperation, p^*, can be altered when a bottleneck of a certain size is imposed and held constant across the whole population (because bottlenecking causes assortment, Section 4.4). While a single round of bottlenecking is expected to have the effect shown in Figure 4.7 above, multiple rounds of bottlenecking (i.e. multiple generations) intercalated with episodes of within-collective selection, mutation and drift will have an effect that is difficult to predict but can be simulated straightforwardly. The behaviour of the model is expected to differ depending on the base game (i.e. the unmodified social niche), determined by model parameters $\left(\begin{smallmatrix} R & S \\ T & P \end{smallmatrix}\right)$.

A population was initialised with $N = 200$ collectives, each with $n^0 = 1024$ particles.

The Bottleneck Size Preference trait's mutation rate was set to zero, so all particles in all collectives would start and remain monomorphic for Bottleneck Size Preference within a simulation run. This is to test the effect of an exogenously imposed bottleneck of a given size on the global equilibrium level of cooperation. (Evolving bottleneck sizes and social niche construction come next, in Section 4.7.)

The base game played between particles within each collective (and determining their unmodified social niche) was varied. The simulation was repeated for all bottleneck sizes for the following base games: Prisoners' Dilemma, Snowdrift game and Stag Hunt game (defector-dominant region). The biological significance of these different games was reviewed in Section 3.8.1 (boxed text on page 45) above.

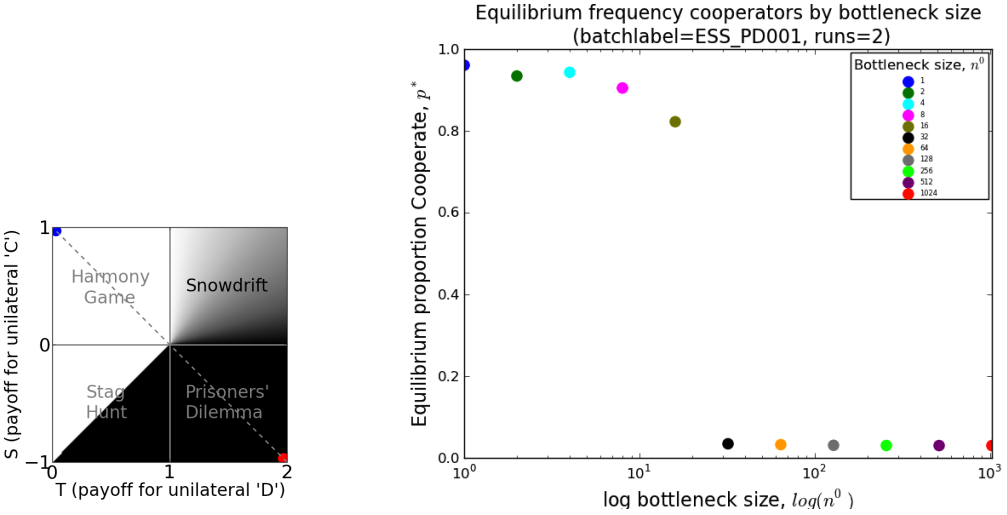

FIGURE 4.10: LEFT: The base game (a simulation parameter) is a Prisoners' Dilemma, marked with a red dot on the T-S plane. We know from prior theoretical considerations that full assortment would result in a Harmony Game (blue dot). We expect (Section 3.8.4) that incrementally increasing assortment from none to full moves the social niche along the dashed line. RIGHT: Results show the equilibrium level of cooperation, p^*, plotted against log bottleneck size, $log(n^0)$. When there is no bottleneck, the social niche is unmodified and the equilibrium frequency of cooperation is near zero (red dot, bottom-right corner). Increasing bottlenecking increases strategy assortment, so changing the game into a more harmonious game. When the social niche is modified enough, the effective game changes from a Prisoners' Dilemma to a Harmony Game suddenly. (All plots in this chapter use the same colour pallet for bottleneck size, to facilitate comparison.)

4.6.1 Results - effect of bottleneck size on equilibrium frequency of cooperation when unmodified social niche is a Prisoners' Dilemma

The population was initialised with all particles as defectors[10]. Figure 4.10 shows the equilibrium level of cooperation for each bottleneck size (exogenously imposed) for a base game of a Prisoners' Dilemma ($S = -1, T = 2$). When the bottleneck size, n^0, is 1024 particles, the equilibrium level of cooperation, p^*, is near[11] zero (red dot). There is no discernible response to mild and intermediate bottlenecks, followed by a sudden change to high levels of cooperation in the middle of the (log) scale. This can be understood by comparing the results scatter plot on the right side of Figure 4.10 with the T-S plane diagram to its left. We know bottlenecking will move the effective game along a straight line on the T-S plane from the base game (PD) toward the upper-left corner, the Harmony Game. (The geometric interpretation of assortment on the T-S plane was explained in Section 3.8.4, page 49). Along this line, there is just such a sudden change in equilibrium level of cooperation (p^*), on moving from the Prisoners' Dilemma region (shaded black, $p^* = 0$) into the Harmony Game region (shaded white, $p^* = 1$) at the centre of the plot. Tighter bottlenecks do not change the equilibrium level of cooperation (p^*) while the

[10]For the Prisoners' Dilemma game, the initial strategy distribution doesn't make any difference to the equilibration behaviour.

[11]Mutation/selection balance means that the theoretical minimum and maximum p^* are never observed, there is always some mutation noise.

effective game remains in the black region. However, once the level of bottlenecking is high enough, the assortment is high enough to move the social niche into the Harmony Game region and the equilibrium level of cooperation suddenly changes. Within the Harmony Game region (shaded white, $p^* = 0$), the equilibrium frequency of cooperation is 1 but the realised frequency is sometimes less than this, due to mutation noise. However, it increases with increasing bottleneck severity. To understand why, recall that the social niche determines the strength and direction of selection on social behaviour (page 33). Under sufficient bottlenecking the social niche is modified such that the *direction* of selection on social behaviour is to favour cooperation but there is still variation in the *strength* of that selection - hence the realised frequency of cooperation increases with increasing bottlenecking even within the Harmony Game region.

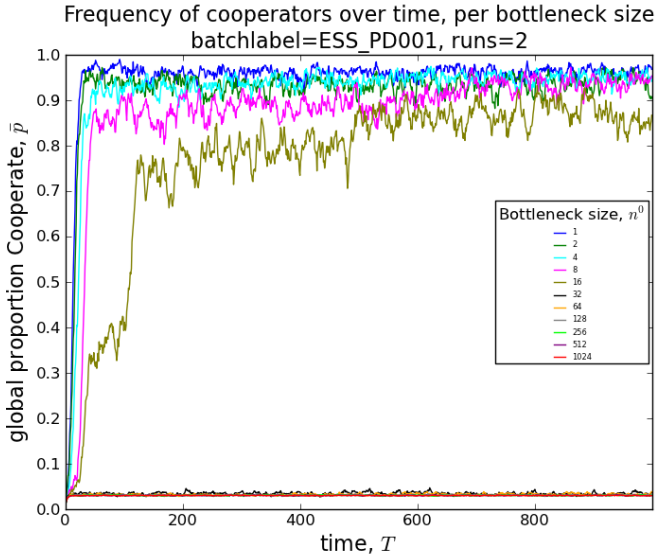

FIGURE 4.11: Simulation result for a typical run: Time evolution of global proportion of cooperators $\bar{p}$, by bottleneck size, n^0, for base game = Prisoners' Dilemma ($T = 2, S = -1$). When running for very much longer (10^5 generations), bottleneck sizes up to 128 can eventually support cooperation for the parameters used here. The numbers are not important - what is important is that the effect does not rely on the formation of pure groups.

Figure 4.11 shows the time-evolution of the frequency of cooperators during the same simulation run. For those bottleneck sizes that will support high levels of cooperation, note how the bottlenecking takes effect very quickly when bottleneck size n^0=1 but increasingly slowly as n^0 is increased. Recall that a social niche (represented as a point on the T-S plane) affects both the direction *and strength* of selection on social behaviour. For those bottleneck sizes that will support high levels of cooperation, smaller bottlenecks are expected to be subject to stronger selection than weaker ones and this is clear in Figure 4.11.

4.6.2 Results - effect of bottleneck size on equilibrium frequency of cooperation when unmodified social niche is a Snowdrift game

We can repeat the above simulation using the Snowdrift game as the base game (in this case $S = 0.25, T = 1.25$, Figure 4.12). When the base game is a Snowdrift, we expect the unmodified

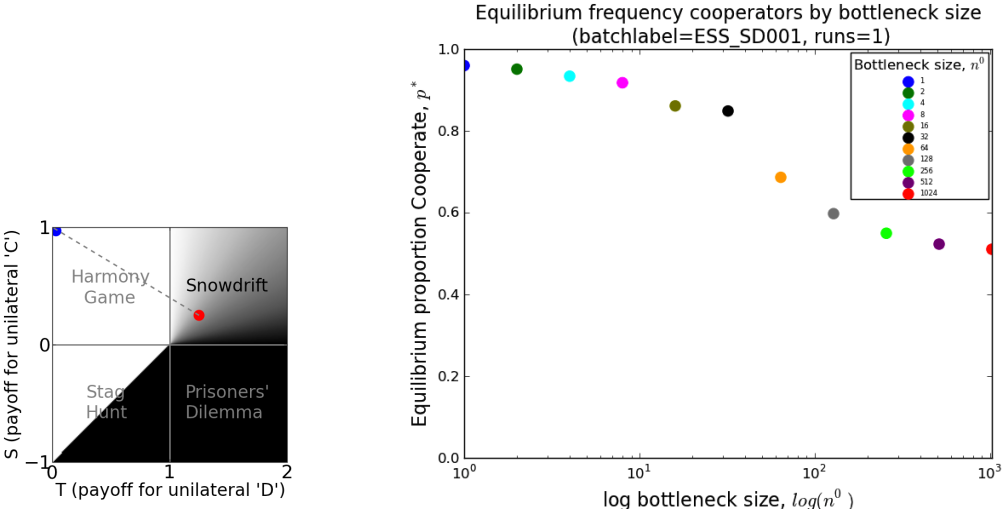

FIGURE 4.12: LEFT: The base game (a simulation parameter) is a Snowdrift game, marked with a red dot on the T-S plane. We know from prior theoretical considerations that full assortment would result in a Harmony Game (blue dot). We expect (Section 3.8.4) that incrementally increasing assortment from none to full moves the social niche along the dashed line. RIGHT: Results show the equilibrium level of cooperation, p^*, plotted against log bottleneck size, $log(n^0)$. When there is no bottleneck, the social niche is unmodified and the equilibrium frequency of cooperation corresponds to the expected mixed-strategy equilibrium of $p^* = 0.5$ (red dot). When bottlenecking is increased, this modifies the social niche by increasing strategy assortment. The results are consistent with the interpretation that the effective game changes gradually from a Snowdrift to a Harmony Game as assortment increases. (All plots in this chapter use the same colour pallet for bottleneck size, to facilitate comparison.)

social niche to support a certain intermediate level of cooperation (this is because, as we saw in Chapter 3, $S > P$ in Snowdrift games, making unilateral cooperation preferable to mutual defection). This mixed strategy equilibrium in a Snowdrift game is given by:

$$p^* = \frac{P - S}{R + P - S - T} \tag{4.6.1}$$

and this is equal to 0.5 for the values of R, S, T and P used to define this Snowdrift game. As expected, we find that in the unmodified social niche (no bottleneck, red dot), the frequency of cooperators equilibrates around the game's mixed-strategy equilibrium of $p^* = 0.5$. As bottlenecking is increased, the equilibrium frequency of cooperators increases. With a one-particle bottleneck (blue dot), cooperation fixates (Figure 4.12, right side).

4.6.3 Results - effect of bottleneck size on equilibrium frequency of cooperation when unmodified social niche is a Stag Hunt game

Stag Hunt games are unusual among the games considered here in that the initial frequency of cooperators (p_0) affects the outcome. Stag Hunts are coordination games with two stable equilibria, one at proportion of cooperators $p = 1$ (called the payoff dominant equilibrium) and another at $p = 0$ (called the risk-dominant equilibrium). They also have an unstable equilibrium at:

$$p^* = \frac{P - S}{R + P - S - T} \tag{4.6.2}$$

The significance of the unstable equilibrium is that it defines the watershed between two basins of attraction, one where cooperation is favoured and one where defection is favoured. Whichever side of the watershed the population starts at will determine its future state.

For Stag Hunt games, the shading on T-S plane diagrams is interpreted differently compared to other games, because the values of R, S, T and P are insufficient to determine the equilibrium frequency of cooperation (p_0 is also needed). In the Stag Hunt case, the angle subtended by the boundary between the black-shaded and white-shaded regions indicates the position of the watershed, p^*. For Stag Hunt games where the defection-dominant basin is at a maximum ($p^* = 0$), the entire Stag Hunt region would be shaded black, while for Stag Hunt games where the defection-dominant basin is at a minimum ($p^* = 1$), the entire Stag Hunt region would be shaded white. In the example presented here ($R = 1, S = -0.75, T = 0.5, P = 0$), $p^* = 0.5$ and so the angle subtended by the watershed is half way between those two extremes. The population was initialised with 100% defectors in each collective (so we expect no cooperation in the unmodified social niche).

Results are presented in Figure 4.13. When there the social niche is unmodified (large bottleneck size) and the initial frequency of cooperators is low, the system stays in the risk-dominant attractor and defection is stable, as expected (red, magenta, green dots). Below some critical bottleneck size (for the parameters used here it is 128 particles, grey dot) there is sufficient assortment to cause the system to move to the payoff-dominant attractor. Assortment is key to coordination games like the Stag Hunt. Skyrms' precis of how populations faced with Stag Hunts come to move from the defector-dominated equilibrium to the cooperator-dominated equilibrium gives an intuitive explanation as follows:

> "How much progress have we made in addressing the fundamental question of the social contract: 'How can you get from the noncooperative hare hunting [defecting] equilibrium to the cooperative stag hunting equilibrium?' The outlines of a general answer have begun to emerge. Over time there is some low level of experimentation with stag hunting [i.e. cooperating]. Eventually a small group of stag hunters comes to interact largely or exclusively with each other. This can come to pass through pure chance and the passage of time in a situation of interaction with neighbors.

> ...The small group of stag hunters prospers and can spread by reproduction or imitation." (Skyrms, 2004, Postscript)

In the present case, the 'low level of experimentation' comes from mutation causing cooperators to come into existence spontaneously from time to time. The 'small group of stag hunters [that] comes to interact largely or exclusively with each other' is a collective with a small enough bottleneck size (and enough luck in the sampling process) to contain an above-average proportion of cooperators. The smaller the bottleneck, the more likely that sampling error would cause some collectives to be composed of an above-average proportion of co-operators (higher assortment) and thus fitter than other collectives, so spreading cooperation in the population. In the example here, bottlenecks of size 128 (grey dot) and below suffice.

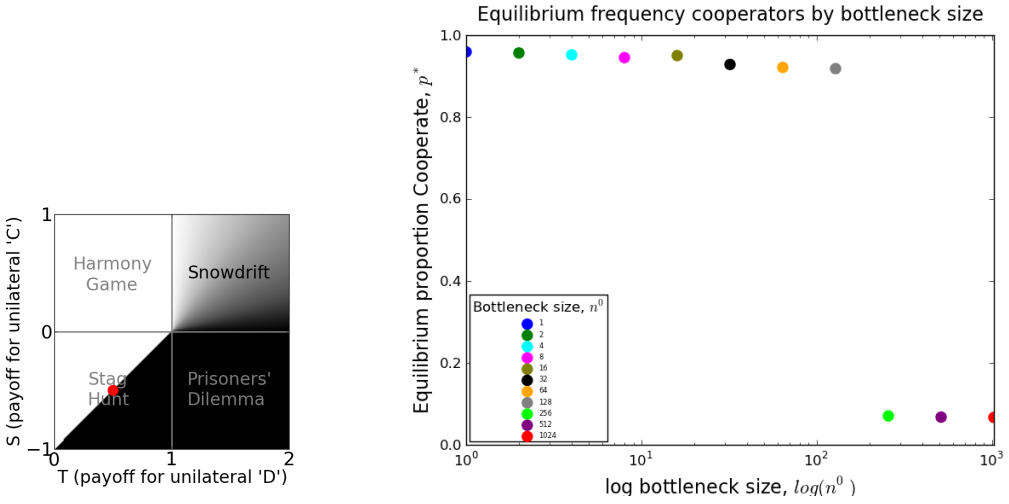

FIGURE 4.13: LEFT: Simulation parameters - the base game is a Stag Hunt, marked with a red dot on the T-S plane. For this Stag Hunt game, an initial frequency of cooperators $p^* > 0.5$ will result in cooperators fixating at equilibrium, while an initial frequency of cooperators $p^* < 0.5$ will result in defectors fixating at equilibrium. The population is initialised entirely with defectors. RIGHT: Results show the equilibrium level of cooperation, p^*, plotted against log bottleneck size, $log(n^0)$ for exogenously imposed bottlenecks from 1024 (red dot) to 1 (blue dot). When there is no bottleneck, the social niche is unmodified and the equilibrium frequency of cooperation is near zero, as expected for the base game and initial proportion of cooperators (red dot). Gradually reducing the bottleneck size does not change the equilibrium frequency of cooperation, as the system remains within the defection-dominant basin of attraction (dots magenta and green). When bottleneck size is reduced below some critical threshold, the induced assortment is sufficient to change the game enough to move into the cooperation-dominant basin of attraction (dots grey to blue). See main text for interpretation. (All plots in this chapter use the same colour pallet for bottleneck size, to facilitate comparison.)

4.6.4 Summary of equilibrium analysis for exogenously fixed bottleneck sizes

This preliminary study has shown that exogenously altering the bottleneck severity uniformly over a population of collectives has the effect of altering the equilibrium frequency of cooperation sustainable within those collectives. The effect is different for different base games but

always present. This shows that the model is functioning as expected when presented with initial conditions and parameters for which we have a known outcome (red dots, for all games). It also confirms that bottlenecking is causing assortment of strategies, as cooperation would not have been stable without it for the Prisoners' Dilemma and Stag Hunt cases and would not have been stable at the observed level for the Snowdrift case.

This gives reassurance that the bottleneck is having the expected effect before going on to perform the main experiment, where social niche construction is the focus.

4.7 Simulation: Bottleneck evolution by Social Niche Construction

When both the Social Trait and the Social Niche Modifier (Bottleneck Size Preference) are allowed to vary, particles become both subject and object of their own social evolution - and social niche construction can occur.

The behaviour of the model is again expected to differ depending on the location of the base game, determined by model parameters $\left(\begin{smallmatrix} R & S \\ T & P \end{smallmatrix}\right)$ on the T-S plane. The base game is the unmodified social niche. Recall that different base games each represent different collective action problems and each of them exists in the biological world in different contexts (see boxed text in Section 3, page 45).

A population was initialised with $N = 200$ collectives, each with $n^0 = 1024$ particles. The mutation rate for the Social Trait was set to 10^{-2}. The mutation rate for the Social Niche Modifying Trait (Bottleneck Size Preference) was set to $5 \times 10^{-}4$. (This is deliberately less than the mutation rate for the Social Trait, so that social behaviour can catch up with and stabilize within the new social niche before the social niche changes again. This separation of timescales is non-essential but makes interpreting the results more straightforward.)

The base game played between particles within each collective (and determining their unmodified social niche) was varied. The simulation was repeated for base games as follows:

- Prisoners' Dilemma game ($S = -1, T = 2$)
- Snowdrift game ($S = 0.25, T = 1.25$)
- Stag Hunt game ($S = -0.5, T = 0.5$ and defector dominant initial strategy frequencies)

The results sections below show the position of the base game (representing the unmodified social niche) on the T-S plane for each of the base games studied.

All particles were initialised with the same large Bottleneck Size Preference trait. The initial distribution of Social Trait among particles was all defectors.

The length of time required for meaningful results to be observed varies considerably with the initial Bottleneck Size Preference and underlying game. As discussed in Section 4.4 above, the effect of bottlenecking on assortment varies as $1/n$ and only has an effect when there is variation

in the parent from whom the sampling is being done. Therefore, at large bottleneck sizes, drift dominates selection and simulations runs can be uneventful. Without loss of generality, initial bottleneck sizes were therefore chosen such that something interesting would happen within a reasonable time period.

4.7.1 SNC results when unmodified social niche is a Prisoners' Dilemma

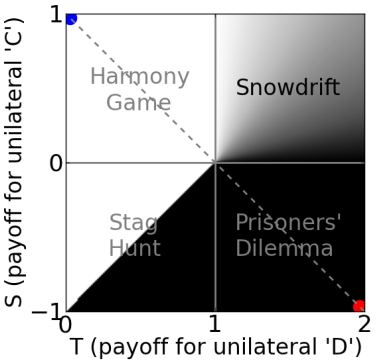

FIGURE 4.14: Simulation parameters: The red dot indicates that the base game used in the simulation is a Prisoners' Dilemma game ($R = 1, S = -1, T = 2, P = 0$). Shading indicates equilibrium frequency of cooperation, p^* (black is 0, white is 1).

Results are displayed in Figure 4.15 for a typical run. (Stochasticity means the timing of interesting dynamics varies from run to run, so a single typical run is shown for illustration.) Cooperation (blue line) is initially disfavoured and bottlenecking (magenta) is weak (128 particles). Cooperation and bottlenecking concurrently evolve, each in response to the other, until very high levels of cooperation and a unitary bottleneck evolves. Higher cooperation leads to higher absolute numbers of particles (green line). Over the same time period, the level at which character-fitness covariance obtains (black line) shifts from particle-level to collective-level. The black line shows an evolutionary transition in the level of individuality, as defined in this thesis.

For a social niche that unmodified would be a Prisoners' Dilemma, we expect a sharp transition in the alignment of particles' fitness interests once the level of assortment rises enough to move the effective game being played into the Harmony Game region of the T-S plane. A sharp transition is indeed observed between bottleneck sizes 64 and 32.

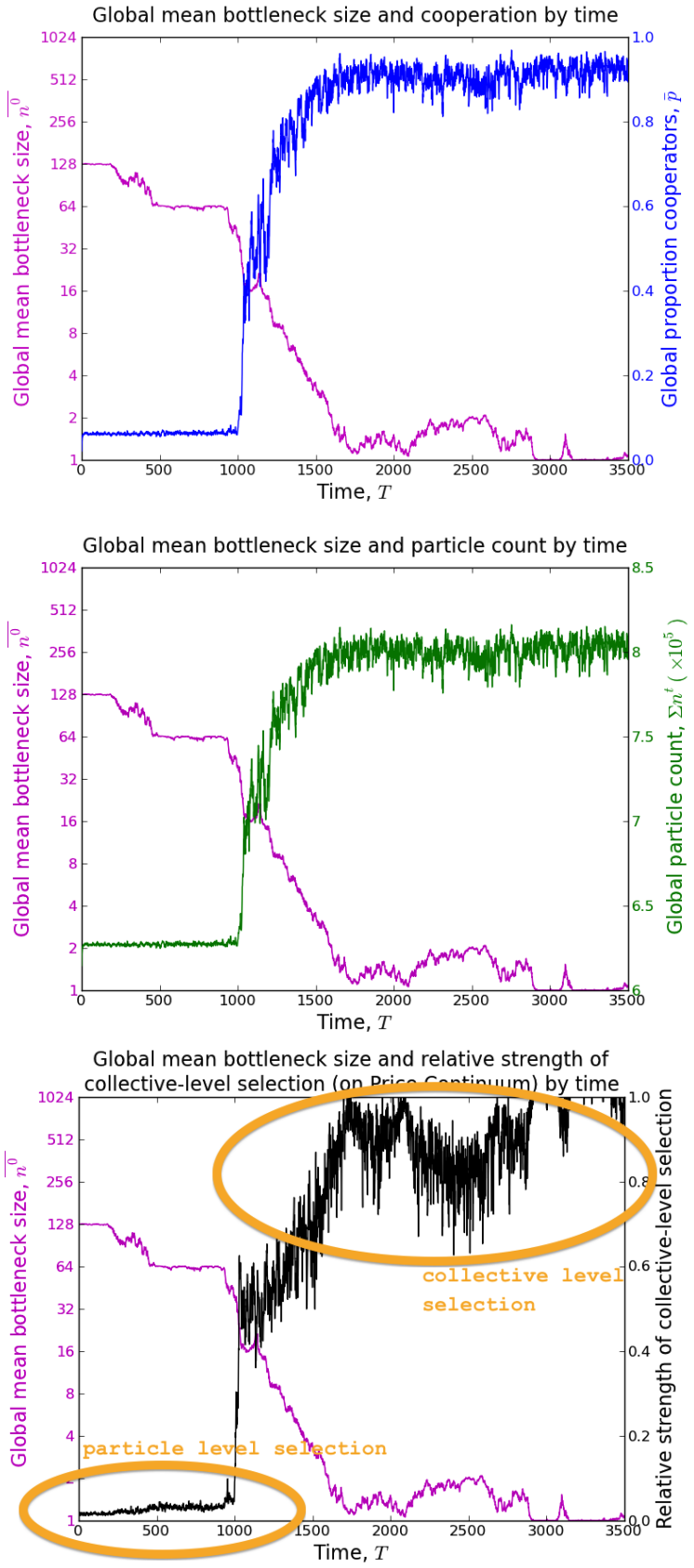

FIGURE 4.15: Time evolution of a typical Prisoners' Dilemma run. Mean bottleneck size (all plots, magenta), global prevalence of cooperation (top plot, blue), global absolute particle number (middle plot, green) and level of individuality as measured by the Price Approach (bottom plot, black) by time. Note how cooperation (blue) is very low initially (held just above zero by mutation noise) but how it increases with decreasing bottleneck size (magenta). Note also how decreasing bottleneck size increases the number of particles sustained in the global population. Social niche construction occurs because tighter bottlenecks sustain more cooperation and more cooperation sustains more particles.

4.7.2 SNC results when unmodified social niche is a Snowdrift game

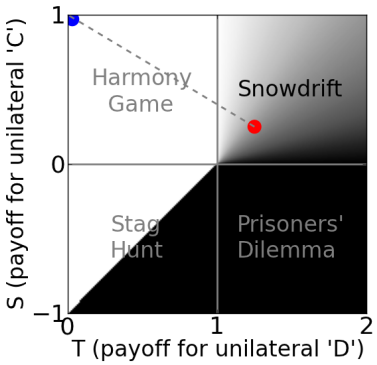

FIGURE 4.16: The red dot indicates that the base game used in the simulation is a Snowdrift game ($R = 1, S = 0.25, T = 1.25, P = 0$). Shading indicates equilibrium frequency of cooperation, p^*.

Results are presented in Figure 4.17. Note that, when the unmodified social niche is a Snowdrift game, an intermediate level of cooperation is sustained regardless of the presence of the Social Niche Modifier (bottleneck). However, higher levels of cooperation result in higher social welfare (and so higher absolute particle number). Bottlenecking and cooperation concurrently evolve in response to each other, until full bottlenecking (a unitary propagule) and near-universal cooperation are reached. (The level of cooperation does not quite reach 100% due to constant mutation noise.) The level of individuality (as measured using the Price Approach) starts intermediate and rises to fully collective-level as bottlenecking increases. Bottlenecking suppresses particle-level selection by denying variance to the particle level. Bottlenecking also increases collective-level heredity. This is because collectives with little internal evolutionary change going on inside them are likely to have a distribution of types at the reproductive stage in their lives that is similar to the distribution they were founded with (Frank, 2012).

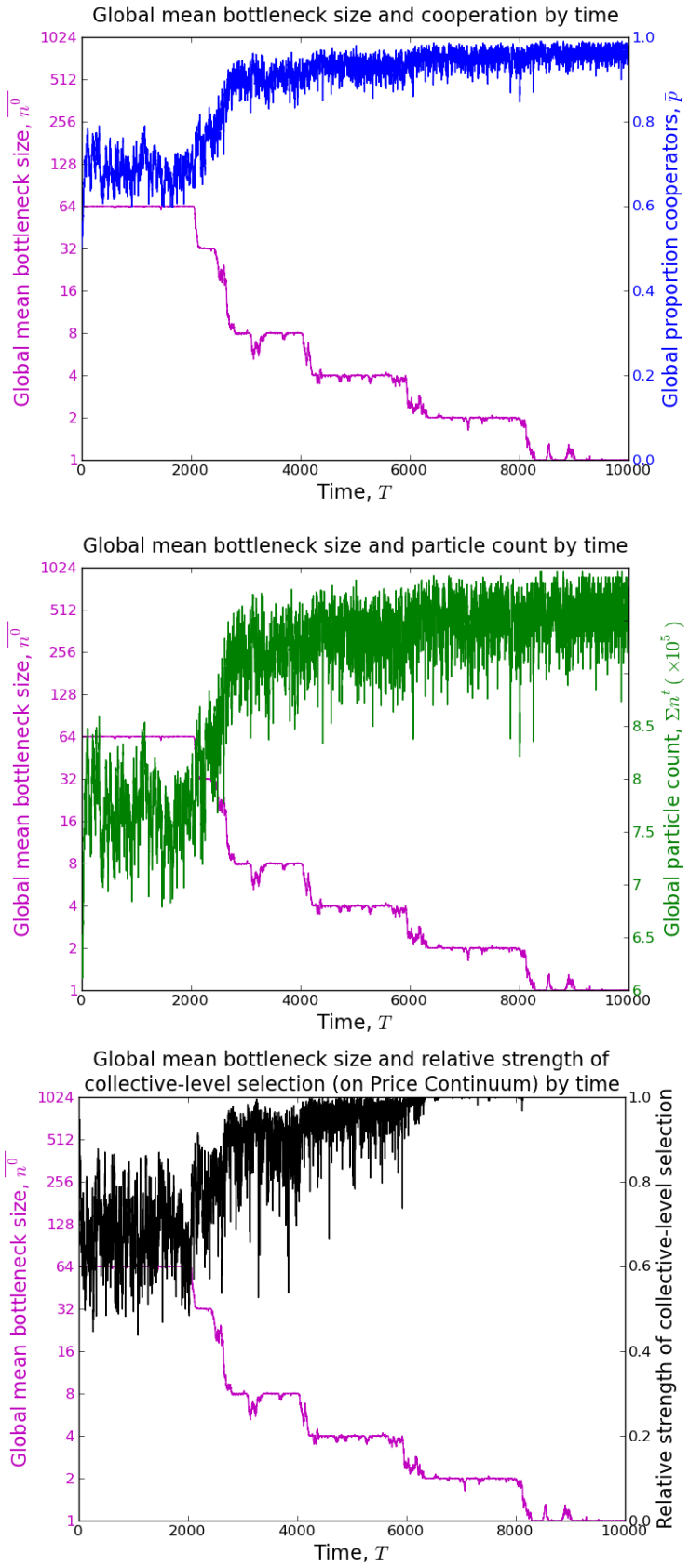

FIGURE 4.17: Time evolution of a typical run, Snowdrift. Mean bottleneck size (all plots, magenta), global prevalence of cooperation (top plot, blue), global absolute particle number (middle plot, green) and level of individuality as measured by the Price Approach (bottom plot, black) by time. Cooperation (blue) starts at mixed-strategy equilibrium for the base game ($p^* = 0.5$) and rises from there to near fixation with increasing bottlenecking. Increasing cooperation sustains larger absolute number of particles (green). The level of individuality (black) starts intermediate and rises to fully collective-level selection (and evolution) with increasing bottlenecking.

4.8 Comparison with other work

The bottleneck model shows social niche construction in action. Social Trait and the Social Niche Modifier reciprocally evolve in response to each other, leading to more cooperative social niches, more cooperation and higher absolute fitness of the particles. (I will return to this point in Chapter 5 in discussing and rejecting the idea that evolutionary transitions involve 'sacrifice for the greater good'.)

Several authors have previously considered the effects of bottlenecking or analogous phenomena on the evolution of cooperation and I now briefly review some of that work and how it relates to the work presented here.

4.8.1 Comparison with Michod and Roze

The present model owes much to (and is partly inspired by) a model due to Michod and Roze (1999; 2001) but there are important differences. Michod and Roze's model is a two-locus modifier model of social evolution (in common with Odling-Smee et al. (2003), Powers et al. (2011), Jackson (2015) and this thesis). Individuals have a locus that represents a social trait and another that represents a modifier (like my 'social niche modifiers') of the conditions of social evolution - here a bottlenecking trait[12]. The overall structure of Michod and Roze's model is similar to mine - a population of particles is organised into collectives; particles play each other in an evolutionary game within collectives (implicitly a Prisoners' Dilemma for Michod and Roze) and particle fitness is proportional to payoff; collectives have a life-cycle involving the emission of propagules of some size (analogous to bottleneck size); collective generations do not overlap.

Michod and Roze investigate the effect on the stability of cooperation of exogenously varying the propagule size, finding that smaller propagule sizes support higher levels of cooperation (analogous to my experiments with exogenously imposed bottleneck sizes on in Section 4.6.1 above). The noteworthy differences concern the modelling assumptions they use, rather than what is, on the face of it, the result of their model.

Michod and Roze do not evolve propagule size (that would enable social niche construction).

Importantly, particle-level and collective-level selection are modelled with different versions of the model, using different functions for collective fitness$_2$ (number of offspring collectives), W. Converting to my notation, Michod and Rozes' two expressions for W are as follows. When there is (on their account) only partial individuality at the higher level:

$$W = (1 + \beta p^t)\frac{n^t}{N} \tag{4.8.1}$$

[12] Michod and Roze (1999) consider modifiers that affect germ-line sequestration (their Section 4.5), mutation rate (Section 4.6), policing (Section 4.7) and - relevant here - bottlenecking (Section 3).

When there is (on their account) true individuality at the higher level:

$$W = 1 + \beta p^t \tag{4.8.2}$$

(β is a measure of the benefit gained from each cooperator, p^t is the proportion of cooperators in the the collective at maturity, n^t is the number of particles in the collective at maturity and N is the bottleneck size in particles.)

Equation 4.8.1 says each collective's fitness$_2$ is a function of both the proportion of cooperators at maturity, p^t and the particle number at maturity, n^t. If β is low (near zero) then particle fitness is the main determinant of collective fitness. Michod and Roze (1999, p.10) say Equation 4.8.1 "underscores the lack of true individuality in these early cell groups, since there is a direct contribution of cell fitness to organism fitness", presupposing that this must not be the case for fully fledged organisms. (Presumably they do not consider fungi, plants or animals that employ somatic embryogenesis 'true individuals' as, for all these species, fitness depends to a large extent on cell number; it is only for Weismannian organisms that this is not the case.)

Equation 4.8.2 says each collective's fitness$_2$ is a linear function of the proportion of cooperators at maturity, p^t. That is, collective fitness$_2$ depends on the *functionality* (defined as cooperativeness) of the adult collective, not on how *fat* it is. There is no component of collective fitness$_2$ here that depends on particle fitness - fitnesses are *decoupled*. This, say Michod and Roze, is the trademark of true evolutionary individuality at collective level,

There are several important things to note here. Firstly, Michod and Roze have built into their model assumptions a distinction between evolutionary processes acting at higher (Equation 4.8.2) and lower (Equation 4.8.1) levels. In my model there is no explicit change of collective fitness function for different stages in the evolution of individuality - my collective fitness always depends on particle number but still individuality (as I have defined it) emerges at collective level. That is, in my model whether heritable variation in fitness is at particle-level or collective-level (or somewhere in-between) is determined by state variables that change during model execution.

Secondly, I differ from Michod and Roze in that I use covariance between entity fitness and offspring character as the determinant of the evolutionary unit while they prefer to focus on the relation between particle fitness and collective fitness (and whether those quantities are 'decoupled' or not). I do not use the distinction between collective fitness$_1$ (collective fitness understood as aggregated particle fitness) and fitness$_2$ (collective fitness in its own right - number of offspring collectives) to explain evolutionary transitions (Okasha, 2006, Table 8.1).

I have a lot more to say about the work of Michod and colleagues more broadly - a topic I return to in the following chapter.

4.8.2 Comparison with Traulsen and Nowak

Traulsen and Nowak (2006) present a model with some similarities to the bottleneck model in this chapter. Their model involves a population of individuals (particles) structured into groups (collectives). Individuals play the others in their group in an evolutionary game, yielding a payoff that is used as a proxy for fitness. The game is, implicitly, a Prisoners' Dilemma - no other games are considered. Individuals are altruists or defectors. Altruistic individuals are at a disadvantage within groups but groups containing more altruists grow large more quickly. The model update algorithm is a Moran process as follows. At each time step, one particle is chosen for reproduction from the global population with probability proportional to its fitness. The new particle is a copy of the pre-existing one and is added to the same group. When groups get to a maximum size (n), they split in half with some probability q (or one individual dies with probability $1 - q$). This means groups are founded by sampling from parent groups using a sample size equal to $n/2$ (this is analogous to my bottleneck size). Each time a new group is founded, another group is selected at random and removed, so maintaining a constant number of groups m.

Trauslen and Nowak experiment with different exogenously imposed values for n, finding that smaller n favours greater cooperation (this is analogous to my experiments with exogenously imposed bottleneck sizes on in Section 4.6.1 above). The driver of the evolutionary dynamics is between-group variance driven by sampling error. As discussed earlier in this chapter, smaller samples reduce variance in offspring groups, so strengthening between-group selection and weakening within-group selection.

This model's behaviour is consistent with the bottleneck model where comparison is possible. There are two major differences: Firstly, the possibility of allowing n to concurrently evolve with strategy is not considered (that would have enabled social niche construction to act). Secondly, the authors (rightly) consider the evolutionary stability of altruism to be symptomatic of higher level selection dominating lower level selection in their model. The effects on group-level heritability are not discussed, nor the possibility of the higher level becoming the level at which the evolutionary process operates.

4.8.3 Comparison with Godfrey-Smith

In Chapter 2 I discussed Godfrey-Smith's (2009) multi-dimensional conceptual framework for understanding what he calls Darwinian Individuality. Recall that an important dimension in his scheme for collectives is dimension B, the extent to which there is a life-history bottleneck. Godfrey-Smith discusses two reasons why bottlenecking contributes to Darwinian Individuality. One is, he suggests, that a new start every generation makes possible variation that affects the organism's basic organisation - an internalist appeal to developmental constraint (or a means of overcoming it to some extent). He also puts forward another explanation, close to that pursued

in this chapter, suggesting that bottlenecking constrains the availability of variation within collectives and increases it between them, so acting to shift Darwinian Individuality to collective level. This is a process he calls the 'de-Darwinization' of the lower level entities and refers, I suggest, to social niche construction (without explaining how it occurs).

> "A bottleneck is a narrowing that marks the divide between generations. This narrowing is often extreme - to a single cell - but in principle is a matter of degree. So the degree of B is the 'degree of bottleneckishness', the extent of the narrowing. This might be understood absolutely, or as a relation between adult and initial size. In the clearest cases we find both. ... The clearest cases where B is high are those where there are zygotes and other one-celled beginnings. ... But B is intended here as gradient matter, not as a distinction which puts one-celled beginnings in one category and everything else in another. ... As John Matthewson pointed out, an absolute measure is more important in this de-Darwinizing role, while a relative measure is probably more important in the earlier role concerning the supply of variation in the evolution of collectives." (Godfrey-Smith, 2009, Sections 5.2 and 5.3)

As discussed in Section 4.4 (page 71) above, it is indeed an absolute rather than a relative bottleneck that is important to the de-Darwinizing effect of bottlenecking. This is because the reduction in sample variance for a small sample depends only on the size of the sample and not the size of the population from which the sample is taken. What Godfrey-Smith does not acknowledge is that only small absolute bottleneck sizes make a difference to this effect. He places considerable emphasis on bottlenecking being a gradient concept yet he fails to discuss the strong non-linearity of the effect. It seems likely, therefore, that explanations for what Godfrey-Smith calls 'intermediate' bottlenecks - those containing very many cells (but nevertheless a small proportion of the adult cell number) - are to be sought elsewhere and are probably not variance/assortment related.

Chapter 5

The Fitness Export Theory for evolutionary transitions in individuality

5.1 Overview of chapter

The goal of this thesis is to offer an evolutionary explanation for how evolutionary transitions in individuality (ETIs) occur. However, there is already a well-known and different theory that attempts to explain the same thing. That is the Fitness Export Theory[1] associated with Michod and colleagues. In this chapter, I start by explaining the appeal of Fitness Export Theory before going on to offer a critique of it. In Chapter 3, I offered a rival evolutionary explanation for how ETIs occur. I contend that my explanation explains what Fitness Export Theory tries to explain but does not suffer the same shortcomings.

The central motivating intuition behind Fitness Export Theory is the notion of self-sacrifice for the greater (collective) good. The idea is that the evolution of altruism among particles causes them to decrease in fitness, while collectives of them increase in fitness. This brings about the export of fitness and so evolutionary individuality from particle to collective level. Once the evolutionary process has started to act at collective level, it is then possible for mechanisms to evolve that suppress within-collective conflict, so consolidating individuality at collective level.

My critique of Fitness Export Theory rests on multiple objections. I find problematic both the self-sacrifice intuition and the idea that fitness-bearing is constitutive of biological individuality. I offer a diagnosis of where Fitness Export Theory has gone wrong and how it has seemed persuasive to many, despite its shortcomings. The following chapter suggests ways forward.

[1] Michod et al. do not explicitly name their theory. I refer to this theory many times, so for ease of reference I have named it Fitness Export Theory in this thesis. It is sometimes known by the alternative names 'fitness decoupling' theory (Okasha, 2006) or 'MVSHN' theory (after the authors of one of the key papers - Bossert et al., 2013).

5.2 Fitness Export Theory as a diachronic explanation for ETIs

Recall that the explanatory challenge (set out in Section 1.4) is framed as follows: Before an ETI gets started, there is single-level heritable fitness variance at the level of particles. Particles here are evolutionary units and their fitness covaries with the character of their offspring. After an ETI is consolidated, there is single-level heritable fitness variance at the level of collectives. Collectives here are evolutionary units. Collectives' fitness now covaries with the character of their offspring - and this has ceased to be true of particles. The challenge is to explain that transition in evolutionary terms. (Quite plausibly, many authors suspect multi-level selection to be involved somewhere between these start and end points.)

Fitness Export Theory is currently the best-known answer to the explanatory challenge outlined above. The theory has been developed over an extended period of time in a series of publications differing in their emphasis and approach but sharing a particular way of thinking about how collectives come to be evolutionary units in their own right (e.g. Michod, 1999; Michod and Roze, 1999; Michod and Nedelcu, 2003; Michod, 2005; Michod et al., 2006; Michod, 2006). The central theoretical commitment is that ETIs occur due to the evolution of altruism among particles, causing particle fitness to decrease so that collective fitness may increase; this is the 'export of fitness' and individuality to the higher level (Michod, 1999, pages xi, 3, 9, 17, 32, 81, 83, 84, 135, 137, 140, 163). The evolution of social behaviour is thus central to the explanation for ETIs offered by Fitness Export Theory. The evolution of social behaviour is also central to the explanation for ETIs offered in this thesis. However, the explanations are quite different, as will become apparent in what follows.

Fitness Export Theory can be understood as making two moves. The first is to explain how fitness comes to be exported to the higher level - this move involves the evolution of altruism among particles causing a decoupling of particle and collective fitnesses. The second move is to explain why fitness being exported brings about a transition in individuality. This involves the evolution of conflict-suppression mechanisms at collective level, enabled by the prior export of fitness to that level. The theory is supported by plentiful verbal argument and also with a detailed model (the life-history tradeoff model), which I describe in Section 5.3 below.

(Important note: The terms 'cooperation' and 'altruism' are treated as synonyms in many papers on Fitness Export Theory, with 'cooperation' being used more frequently and explicitly defined as *'An interaction that decreases the fitness of the individual but increases the fitness of the group'* (Michod, 1999, p.17). For consistency with other parts of this thesis (Chapter 3) and common usage (e.g. West et al., 2007b) and to avoid confusion with mutually-benefical cooperation, I will use the term 'altruism' to describe an interaction that decreases the fitness of the actor but increases the fitness of the recipient. I have left Michod's own words intact when quoted directly but it must be remembered that his definition of 'cooperation' is what I call altruism. Nothing I say in this chapter turns on an altruism/cooperation distinction, this is simply an attempt to maintain consistent terminology throughout the thesis.)

Consider first the role of the evolution of altruism in Fitness Export Theory.

> "Because cooperation reduces the fitness of lower level units, while increasing the fitness of the group, cooperation drives transitions to higher level units." (Michod, 1999, p.3)

> "Cooperation can produce a new more complex evolutionary unit, with the requisite properties of heritable fitness variations, because cooperation trades fitness from a lower-level (the costs of cooperation) to the higher-level (the benefits for the group)." (Michod et al., 2003)

> "The evolution of cooperation is fundamental to ETIs, because it exports fitness from the lower level (e.g., its costs to cells) to the higher level (its benefits to the group) and in this way cooperation may create new levels of fitness." (Michod, 2005)

> "The evolution of cooperation, the central problem of social biology, gains special significance during ETIs because altruism and other forms of cooperation lead to the transfer of fitness from the lower level (the costs of altruism) to the group level (the benefits of altruism)." (Hanschen, Shelton and Michod, 2015)

Fitness Export Theory recognises that the collective living involved in evolutionary transitions entails a social dilemma (as I have done in this thesis, owing much to Michod, 1999). It adopts the vocabulary of the game-theoretic approach to social evolution, with talk of altruistic acts incurring a 'cost' for the actor and a 'benefit' for the recipient. This talk of fitness costs and benefits is load-bearing in Fitness Export Theory's account of how fitness gets exported from the particle level to the collective level. A group-structured population of particles and an MLS1-type group-selection process is assumed ('MLS1' was defined in Section 2.4). Altruism increases in frequency in the global particle population because, it is further assumed, collectives containing more altruists do better (i.e. contribute more particles to future generations), even though altruists are at a disadvantage within collectives. The export of fitness is due to the nature of altruism, where altruistic acts confer a fitness benefit upon the recipient (the collective) at a fitness cost to the actor (the particle). This means that, as a direct result of the evolution of altruism, overall particle fitness is reduced while collective fitness is increased. Eventually, the evolution of altruism proceeds to the extent that the particles are left with zero fitness and all the fitness present is collective-level fitness. This fitness sacrifice on the part of the particles amounts to the very property of fitness being 'exported' to the collective level, with particles ceasing to be fitness bearers at all.

The next move Fitness Export Theory makes is to argue that the export of fitness amounts to the export of individuality. Fitness Export theory places a lot of emphasis on fitness-bearing and

the MLS1/MLS2 distinction. To see how, consider Okasha's (2006) summary of the process, reproduced in Table 5.1, giving an idealised diachronic model for ETIs as passing though a number of stages corresponding to different selection regimes[2].

[Stage 0]	[SLS]	[particles are units of evolution, particle fitness is unproblematic]
Stage 1	[MLS1]	collective fitness defined as average particle fitness [i.e. collective fitness$_1$] (cooperation spreads among particles)
Stage 2	[grey area]	collective fitness is not defined as average particle fitness but is proportional to average particle fitness (collectives start to emerge as entities in their own right with life cycles of their own)
Stage 3	[MLS2]	collective fitness [i.e. collective fitness$_2$] is neither defined as nor proportional to average particle fitness (collectives have fully emerged; fitnesses are decoupled)
[Stage 4]	[SLS]	[collectives are units of evolution, collective fitness$_2$ is the only fitness there is, particle fitness has gone]

TABLE 5.1: Adapted from Okasha (2006, Table 8.1) 'Relation between collective fitness and particle fitness during an evolutionary transition'. For completeness, I have added a Stage 0, representing single-level selection at particle level prior to the onset of a transition and a Stage 4, representing single-level selection at collective level after completion of a transition. My additions are shown in square brackets.

The narrative starts with an SLS (Single-Level Selection) regime, featuring a population of particles as evolutionary units (Stage 0 in Table 5.1). Only particle fitness exists and it is conceptually unproblematic (or, rather, no more problematic than fitness is outside the context of evolutionary transitions). The process then passes through an MLS1 phase (Okasha's Stage 1 in Table 5.1) where both particles and collectives have fitness (but collective fitness here is defined as an aggregate of particle fitness: collective fitness$_1$). Next it passes through a phase (Stage 2) where, due to the evolution of altruism, collective fitness starts to decouple from particle fitness and to become a new quantity in the world (collective fitness$_2$). There follows an MLS2 phase (Stage 3) where both particles and collectives have fitnesses in their own right (collective fitness here is collective fitness$_2$, now fully decoupled from particle fitness). Then the strength of collective-level selection in that MLS2 regime becomes so strong that it dominates particle-level selection, leaving collectives as the new evolutionary units. Thus we arrive at another Single-Level Selection regime, with collectives left as the only fitness-bearers (Stage 4 in Table 5.1). This puts us back where we started, albeit frame-shifted up a level of biological complexity.

If this theory is right, then somewhere inside the 'grey-area' between Stages 1 and 2, collective fitness$_2$ comes into being for the first time as an essential part of the transition (Michod, 2005). Michod and Nedelcu (2003, p.66) provide a summary of what happens to fitness during ETIs:

[2]Importantly, this is an idealization. There is no implication that this process always runs to completion, always runs in the same direction or always runs at all. For example, see Herron and Michod (2008) and Herron et al. (2009) for a phylogenetic study of the comings and goings of traits thought to be associated with ETIs in the volvocine algae.

> "Group fitness is, initially, taken to be the average of the lower-level individual fitnesses [i.e. fitness_1]; but as the evolutionary transition proceeds, group fitness becomes decoupled from the fitness of its lower-level components. Indeed, the essence of an evolutionary transition in individuality is that the lower-level individuals must 'relinquish' their 'claim' to individual fitness in favor of the survival and reproduction of the new higher-level unit."

Michod and Roze (1999, p.10) emphasize the difference between Stages 1 and 3 in an interesting MLS model (first mentioned in Chapter 4, Section 4.8.1). They explicitly use different formulae to calculate collective fitness_2 depending on the selection regime they envisage being appropriate to the target of the model. When the model is supposed to represent 'simple organisms on the threshold of multicellular life', their collective fitness_2 is an increasing function of both cellular fitness and the proportion of cooperators within a collective. When the model is supposed to represent multicellular organisms with genuine individuality, their collective fitness_2 is an increasing function of the proportion of cooperators within a collective and is not a function of cellular fitnesses at all. Different models are used for different stages in the transition (unlike the Bottleneck Model I presented in Chapter 4, where the same model exhibits both particle-level and collective-level evolution with different state variables).

After (and only after) fitness export has occurred and fitnesses are decoupled, a collective-level evolutionary process is then able to 'legitimize the new unit once and for all' (Michod, 1999, p.42) by evolving conflict-mediation mechanisms that suppress within-collective selfishness (such as bottlenecks and germ-soma separation). In summary, the process is one of the evolution of altruism, leading to fitness export, yielding a collective-level selection regime, leading to evolution of collective-level adaptations that are conflict-mediators[3] that legitimize the new unit once and for all.

I now consider a well-known model used to illustrate Fitness Export Theory.

5.3 The life-history tradeoff model as an illustration of fitness export

The life-history tradeoff model (Michod, 2005; Michod et al., 2006; Michod, 2006) is presented as a concrete example of fitness being exported to the higher level by the evolution of altruism.

The model is inspired by the natural history of the volvocine algae, a taxon containing species taken to be examples of different points on a continuum between uni-cellularity and complex multi-cellularity (Kirk, 2005).

[3] My 'social niche modifiers' are essentially the same concept as conflict-mediators here. The distinction is that I do not claim cooperation somehow evolves first and then conflict-mediation comes as a consequence of the change in the level of evolutionary unit - I hold that the two co-evolve with one another, each causing the other (Chapter 3).

> "The fitness of any evolutionary unit can be understood in terms of its two basic components: fecundity (reproduction) and viability (survival). By specializing in these essential components, cells relinquish their autonomy in favor of the group; as a result, fitness and individuality are transferred from the level of the cell to the level of the group. The cell group, by virtue of the specialization of its member cells, becomes integrated and indivisible and, hence, a true individual. The evolution of cooperation is fundamental to this process because cooperation exports fitness from the lower level (e.g., its costs to cells) to the higher level (its benefits to the cell group) and in this way may create new levels of fitness." (Michod, 2006)

The justification for the model starts by noting that the fitness of any biological entity can be understood as composed of two basic components - its fecundity (reproductive success) and its viability (survival to reproductive age). Single-celled organisms usually separate viability and fecundity functions temporally, with each occurring in different periods of their life-history (because typically they cannot be performed simultaneously). Multi-cellular organisms, on the other hand, can perform viability and fecundity functions simultaneously by separating them spatially, in different parts of the organism (Michod, 2005).

The basic intuition motivating the model is that in multicellular individuals (such as Volvox Carteri) some cells specialise in reproductive functions (germ cells) while others specialise in viability functions (somatic cells) and that this situation leaves all individual cells with zero particle fitness (because either one or the other of the two basic factors of fitness is zero) while the multicellular organism of which they are a part can have quite high fitness (because it has achieved a division of labour, with each of the two basic components of fitness being simultaneously performed by different parts of itself). This is interpreted as a case of biological altruism. The multi-cellular way of life involves cells sacrificing their own fitness for the fitness of the multi-cellular organism of which they are a part. The transition from a single-celled life-history to a multi-cellular one is then a case of the evolution of altruism. Furthermore, once germ-soma separation has been achieved, the fitness of the higher level collectives is not simply the aggregate of the fitnesses of the component particles (as it would be for collectives lacking true individuality) but is a function of the division of labour achieved by the collective organism as a whole (a hallmark of true individuality: particle and collective fitnesses have been de-coupled).

The details of the model (based on Michod, 2006, PNAS) are as follows. A population of particles ('cells' in Michod's terminology) is structured into collectives ('groups') of N particles.

- The particle fitness, w_i, of an individual cell i is the product of its viability, v_i and its fecundity, b_i.

$$w_i = v_i b_i \tag{5.3.1}$$

- The group fitness, W, of a group of N cells is found by taking the mean of all the cellular viabilities, taking the mean of all the cellular fecundities and then taking the product of

the two means.

$$W = \left(\sum_{i=1}^{N} v_i\right)\left(\sum_{i=1}^{N} b_i\right) = \overline{\mathbf{v}}.\overline{\mathbf{b}} \tag{5.3.2}$$

> "Once the [germ-soma] specialization is complete and the lower level units are specialized in one of the two major fitness components (viability or fecundity), they have no fitness by themselves and so group fitness in the sense of MLS1 [collective fitness$_1$] is null, while group fitness in the sense of MLS2 [collective fitness$_2$] may be quite high."
> (Michod, 2005)

The model has an analytic result and is supported with a simulation model. The analytic result is that, for any group, the group's fitness W exceeds mean particle fitness $\overline{\mathbf{w}}$ by an amount equal to the negative covariance of $\mathbf{v}$ and $\mathbf{b}$ (where $\mathbf{v}$, $\mathbf{b}$ and $\mathbf{w}$ are vectors of the group's cellular viabilities, fecundities and fitnesses respectively):

$$W = \bar{w} - Cov(\mathbf{v}, \mathbf{b}) \tag{5.3.3}$$

Michod (2006) calls Equation (5.3.3) the 'group covariance effect'. Group fitness, W, is maximised when individual-level fecundity, b_i, and viability, v_i, are maximally different for each cell i. In the model, this state of maximum group fitness represents full germ-soma separation. In such a state, the fitness of each particle, $w_i = v_i b_i$ is zero because either v_i or b_i will be zero. Note how this connects with the evolution-of-altruism theme discussed above - the actor (the cell) pays a cost and the recipient (the group) receives a benefit.

The group covariance effect is an analytic consequence of the model assumptions (Equation 5.3.2). This can be shown by working backwards, as follows:

$$W = \bar{w} - Cov(\mathbf{v}, \mathbf{b})$$

$$W = \frac{1}{N}\sum_{i=1}^{N} v_i b_i - Cov(\mathbf{v}, \mathbf{b})$$

Substituting in the 'mean of the products minus the product of the means' definition of covariance:

$$W = \frac{1}{N}\sum_{i=1}^{N} v_i b_i - \left(\frac{1}{N}\sum_{i=1}^{N} v_i b_i - \frac{1}{N}\sum_{i=1}^{N} v_i \frac{1}{N}\sum_{i=1}^{N} b_i\right)$$

We arrive back at the definition of group fitness built into the model by Equation (5.3.2):

$$W = \frac{1}{N}\sum_{i=1}^{N} v_i \frac{1}{N}\sum_{i=1}^{N} b_i$$

So the model's analytic result, the group covariance effect, follows directly from the model's assumptions in Equations (5.3.1) and (5.3.2).

The group covariance effect is tested with an evolutionary simulation model based on the above assumptions (Michod, 2006, PNAS). The simulation model details are as follows:

- A population of cells is structured into $M = 1666$ groups (number of groups held constant)
- Each group (indexed by k) contains $N = 6$ cells (number of cells held constant; cells indexed by i)
- Each cell's fecundity b_i is initially assigned a random real number in the range $0 - 1$
- Once each cell's fecundity b_i is assigned, that cell's viability v_i is then determined by a fecundity-viability tradeoff function, as follows:

$$v_i = e^{-4b_i} \tag{5.3.4}$$

  Importantly, this constraint enforces a *convex* tradeoff between particle viability v_i and particle fecundity b_i (Figure 5.1). Any convex tradeoff function could have been used (Michod et al., 2006, JTB).
- Group fitness for each group (W_k) is calculated according to Equation 5.3.2 ($W = \overline{\mathbf{v}}.\overline{\mathbf{b}}$).
- The next generation's M groups are each formed by sampling N cells from the current generation (with replacement) with a probability proportional to the fitness, W, of the group containing them.
- Generations do not overlap.
- The process is repeated for a number of generations (100).
- The interesting variables to track are the global mean particle fitness ($\overline{w}$) and mean collective fitness ($\overline{W}$). If fitness export is occurring, we expect $\overline{w}$ to approach its minimum and $\overline{W}$ to approach its maximum (given the constraints).

I have re-implemented the life-history tradeoff model as it is specified in the PNAS paper (Michod 2006) and reproduced the results (see Figure 5.2). As the simulation proceeds, mean particle fitness does indeed ($\overline{w}$) tend toward its minimum while mean collective fitness ($\overline{W}$) increases towards its maximum (given the constraints), as expected[4]. The results do not vary with the initial distributions in $\mathbf{b}$ (and hence $\mathbf{v}$) - all initial distributions lead to the same outcome. Also, the results do not depend on the specific tradeoff function used (Equation 5.3.4) - any convex tradeoff would produce similar results.

To conclude, the life-history tradeoff model is taken to illustrate the evolution of altruism causing fitness to be exported from the particle to the collective level. A selection process causing the

[4]The minimum value of $\overline{w}$ is not quite zero because the fecundity/viability tradeoff function (Equation 5.3.4) is not completely symmetrical - see Figure 5.1. (When $b = 1$, v is not quite zero.)

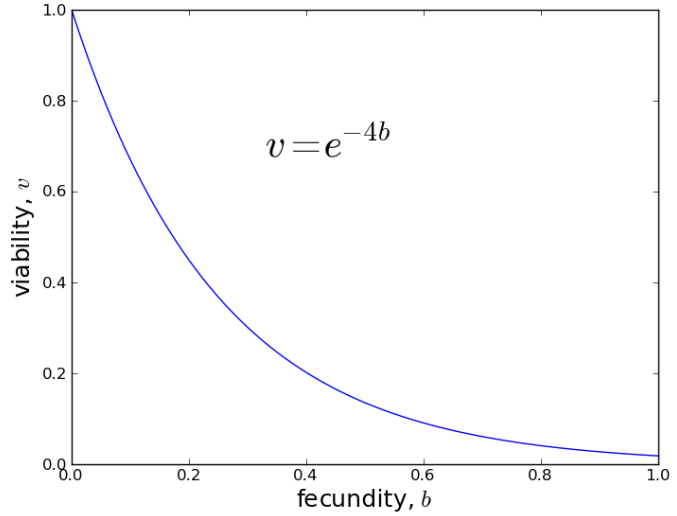

FIGURE 5.1: Michod's (2006) life-history tradeoff model employs a convex tradeoff function constraining the pair of values viability v and fecundity b can take on for any given particle. The function used in the 2006 PNAS paper is $v_i = e^{-4b_i}$, shown here. The 2006 JTB paper makes clear that any convex tradeoff function yields the same results insofar as the 'group covariance effect' is concerned.

mean group fitness to increase also causes the global mean particle fitness to tend towards a minimum. Group fitness is neither defined as nor proportional to aggregate particle fitness - higher and lower level fitnesses are decoupled.

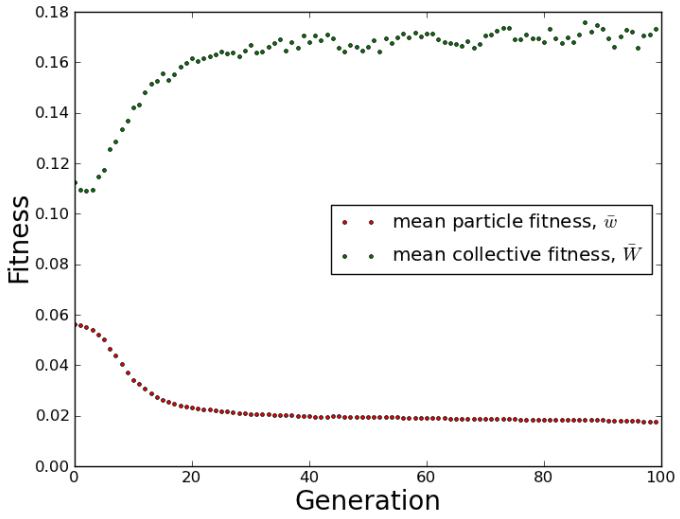

FIGURE 5.2: Simulation results from the life-history tradeoff model, reproducing those of Michod (2006, PNAS, Figure 2). As the simulation proceeds, mean particle fitness ($\bar{w}$) tends toward zero while mean collective fitness ($\bar{W}$) increases towards its theoretical maximum. This is taken to illustrate the evolution of altruism causing fitness to be exported from the particle to the collective level.

5.4 Criticism of evolution-of-altruism argument

This section will be a critique of Fitness Export theory but is not entirely negative. Out of criticism arises a new view on fitness and ETIs that I believe is constructive. I start with the evolution-of-altruism motif, before turning to the life-history-tradeoff model.

The evolution-of-altruism motif recurs in all versions of Fitness Export theory and is central to the explanation it offers. The idea is that, as particle-level altruism evolves (due to collectives with more altruists doing better, even though altruists do worse within collectives), particles pay a fitness cost and the collectives receive a fitness benefit, to the point where the particles eventually have no fitness and the collectives have all of it - and that is how the very property of fitness gets exported from particle to collective level.

In Fitness Export Theory, pro-social behaviour (usually called 'cooperation' in Michod's writings but more properly called 'altruism' for reasons explained above) is defined as 'An interaction that decreases the fitness of the individual but increases the fitness of the group' (Michod, 1999, p.17). This definition does a lot of work in Fitness Export theory because altruism is the very mechanism through which fitness is supposed to be transferred between levels - but it is multiply problematic.

Logical mistake

By convention in evolutionary game theory that makes use of two-player two-strategy symmetric games, interactions may involve a transfer of fitness between an actor and a recipient that are of like kinds. For example, we may write out a payoff matrix such as the following:

		Recipient	
		C	D
Actor	C	$b-c$	$-c$
	D	b	0

Assuming $b > c > 0$, playing 'C' in a one-shot game like this is altruism. According to the definition supplied, the recipient is 'the group'. This means EITHER:

- the game is being played between unlike-kinds (i.e. between an individual and the group of which it is a part, so that one of the players is part of the other one)
  OR
- the game is being played between like-kinds and the group outcome is an aggregate of pairwise games between the individuals within it

The former interpretation is difficult to make sense of, especially when one considers the dimensions of the units in which payoff is measured - collective and particle fitness are measured in different units in Fitness Export Theory. The latter interpretation is more conventional but means that group fitness is understood as an aggregate of individual fitnesses, which is the very MLS1 (strictly, fitness_1) measure of collective fitness that Michod wants to move away from.

Altruism is not costly when reciprocated, which it is when favoured by group selection
There is a further worry with the evolution-of-altruism motif. By definition, an altruistic act (playing 'C' in the above game) is costly to the actor if it is met with defection. However, it need not always be costly to the actor because it need not always be met with defection. In Fitness Export Theory, altruism is supposed to evolve due to between-group selection: groups containing more altruists do better, even though altruists are at a disadvantage within groups. This familiar story relies on the *positive assortment of altruists* within groups to explain how altruism can be favoured more by between-group selection than it is disfavoured by within-group selection. That is, altruism is favoured when it is favoured because it is reciprocated. There is no fitness cost implied by being an altruistic particle in a scenario where altruism is favoured by selection (which is precisely the scenario invoked by Fitness Export Theory). Another way to think about this is to consider the fact that Michod is mainly interested in the formation of clonal groups founded by a single cell, as in the evolution of multicellularity in the volvocine algae. In these situations, close kinship ensures strategy assortment and altruism is beneficial rather than costly to particles. (Recall the discussion in Section 3.8.3 of how the introduction of strategy-assortment modifies games, whether due to relatedness or otherwise).

Objection to the move from 'fitness export' to 'change in level of evolutionary unit'
I differ from Michod and colleagues in that I use covariance between entity fitness and offspring character as the determinant of the evolutionary unit (Sections 1.4 and 2.3.3) while they prefer to focus on the property of fitness-bearing. (See also Okasha, 2006, Table 8.1.) I will argue in Chapter 6 that the bearing of fitness is not the same thing as being a unit of evolution; that discussion is postponed until then. Similarly, the idea that a definition of collective fitness that is influenced by particle fitness (e.g. collective fitness$_1$) is symptomatic of a paucity of collective-level individuality is also found problematic in Chapter 6.

Objection to overall causal story of Fitness Export Theory
The overall narrative of fitness export theory starts with a pre-supposed MLS1 group selection process, where the particle character of interest is altruism[5]. Altruism does indeed evolve and after it has done so (and concomitant fitness export has occurred), the evolutionary process is shifted upward to collective level. Then, Fitness Export Theory has it, this collective-level evolutionary process leads to the evolution of collective-level adaptations such as bottlenecks, germline segregations and policing mechanisms that suppress particle-level selection and 'legitimize the new unit once and for all' (Michod, 1999, p.42).

What I want to object to here is the apparently arbitrary separation between factors supporting cooperation that can reasonably be assumed and factors supporting cooperation that are taken to be explained by Fitness Export Theory. It is assumed that altruism evolves in the system (and this is load-bearing - fitness is exported because altruism evolves). As is well known (Sober and Wilson, 1998, and Section 3.8.3), the evolution of altruism in such models requires[6]

[5]E.g. Michod (2011, p.173): "Cooperation is ... fundamental to ETIs because ... altruism and other forms of cooperation lead to the transfer of fitness from the lower level (the costs of altruism) to the group level (the benefits of altruism). Thus, the evolution of cooperation is the first stage in an ETI."

[6]Fitness Export Theory implicitly assumes the game is a Prisoners' Dilemma.

positive assortment of altruistic behaviours. Only when there is sufficient assortment of social behaviour will between-group selection for altruism outweigh within-group selection against it. Fitness Export Theory implicitly assumes that this positive assortment is present, treating it as exogenous to the explanation offered for ETIs. So far, this follows a fairly standard pattern for an externalist explanation of the evolution of cooperation (Section 3.2). As discussed at length in Chapter 3, strategy-assortment is an example of what I have called a social niche modifier. It is one of those 'external factors' explaining cooperation that is itself usually left unexplained in externalist explanations for social evolution. The origin of the positive assortment of altruists in the MLS1 group selection regime is not addressed in Fitness Export Theory, yet it stands in need of explanation just as much as any other social niche modifier would. Indeed, it stands in need of explanation just as much as the evolution of the collective-level adaptations for conflict-suppression that Fitness Export Theory takes to be part of its explanandum. Even if we accept the claim that fitness is exported from particles to collectives and even if we accept that this export of fitness means export of individuality, there still remains a problem of explanatory postponement.

Fitness export is intended to act as a sort of bootstrapping mechanism that enables escape from the paradoxical situation wherein collectives only seem to exist as evolutionary units in virtue of properties that look like they could only have evolved by a collective-level evolutionary process. Yet part of the explanation offered relies on positive assortment of social behaviour, which itself needs to be explained. There does not seem to be a principled basis for the distinction between the social niche modifiers that can be assumed and those that stand in need of explanation. (By contrast, in social niche construction theory, the social niche modifiers evolve concurrently with social behaviour, each being an evolutionary response to the other.)

5.5 Criticism of life-history-tradeoff model

The verbal argument for Fitness Export Theory is presented semi-informally and perhaps not intended to be interpreted technically. As Godfrey-Smith (2011b) suggests, "this talk of [fitness] export and transfer seems rather metaphorical". However, the idea that particles lose fitness and collectives gain it during an ETI underpins the whole of Fitness Export theory (whether this is due to the evolution of altruism or not). Furthermore, the idea that fitness being exported to the higher level automatically entails an evolutionary transition in individuality is also central to all versions of Fitness Export Theory. Unlike the verbal argument for Fitness Export Theory, the life-history-tradeoff model is precisely constructed and amenable to detailed analysis.

The life-history-tradeoff model is presented as an illustration of how the evolution of altruism at particle level can lead to reduced particle fitness, increased collective fitness and thus export of fitness from the lower to the higher level. It features a de-coupling of particle and collective fitnesses - a hallmark of an ETI accrording to Fitness Export Theory.

It has the following structure.

- If each particle's 'fitness' is the product of its viability and its fecundity (Equation 5.3.1, $w_i = v_i b_i$)
- and if collective fitness is the product of the mean particle viability and the mean particle fecundity (Equation 5.3.2, $W = \bar{v}\bar{b}$)
- and if the relationship between v_i and b_i is convex for each particle (e.g. Equation 5.3.4, $v_i = e^{-4b_i}$)
- then W will be a maximum when particles specialise in either all v and no b or vice versa (and specialisation is altruistic because it causes particle fitness w_i to go down).

De-coupled fitnesses built-in and not explained

There is no doubting the claims of the life-history tradeoff model if it is treated as an exercise in optimizing a function in two variables. However, it is not clear that it is has the explanatory power it is presented as having. It is a matter of mathematical truth that, for a pair of arrays of numbers, the sum of the products is not generally equal to the product of the sums. To define the relation between individual and group fitness in the above way pre-supposes that fitnesses are de-coupled and does not explain how they come to be so. What makes W fail to be an aggregate of w is that the model takes the product of the sum not the sum of the product. If the sum of the product had been used for W, it would be an aggregate of w (i.e. fitness$_1$). Labelling the one 'cell fitness' and the other 'group fitness' does not provide an explanation for how group fitness comes to be decoupled from particle fitness. The execution of the model does not show a move from one expression for W to another - the thing to be explained is built into the model assumptions.

Germ-soma specialisation built-in and not explained

It is an artefact of the convexity of the tradeoff function that W is maximised when each particle specialises in v or b. (This can be seen intuitively by considering Figure 5.1 and a group of just two particles. It is obvious that $(v_1 + v_2) * (b_1 + b_2)$ is maximised when the two particles are placed at the extreme ends of the tradeoff.) As Michod (2006) says, if the tradeoff is linear then all distributions of b (and hence v) produce the same W (i.e. drift occurs) and if the tradeoff is concave then W is maximised when each particle has intermediate values of b (and hence v). So it is an artefact of the model assumptions that W gets higher as w gets lower - that does not tell us anything about the biological world unless there is an isomorphism between the theory and the world - and the case for that is not strong.

Particle viability, fecundity and fitness are causally inert labels

Labelling one of the particle properties 'viability' and the other 'fecundity' and their product 'particle fitness' may invoke a sense of familiarity - many well-known models in theoretical biology break fitness down into the these components (e.g. Gillespie, 1977). However, these variables have no causal power in the model (excluding the causal power they have through their effect on so-called 'group fitness'). The variable labelled 'particle fitness' w in the model has no causal effect on the frequencies of particles within groups within generations. There

is no sense in which particles with higher w get more representation in anything, compared to particles with lower w. Normally in such a model, we would expect a variable labelled 'particle fitness' to *do* something. The model is reminiscent of MLS models where changes in the frequencies of types occur due to selection acting both within and between groups, often in opposite directions. However, only between-group selection operates in this model. There is no within-group within-generation change, so to label variable w as 'particle fitness' is misleading.

Other objections

There is then a further worry - the idea that a quantity going to zero means the quantity ceases to exist. This is not generally true.

Summing up

The life-history tradeoff model does not explain what it purports to explain. The strategy of model-based science is to abstract away contingent detail and to focus on a single phenomenon and its causal structure. The model is known not to correspond exactly to the actual world but does have a causally transparent mechanism. Insight is then gained due to similarity relations between the model and the actual world (Godfrey-Smith, 2006). I doubt that a suitable similarity relation holds here. With Godfrey-Smith (2011a), I don't accept that the cells in a multicellular organism have zero fitness. I will grant that they may not have *heritable variance in* fitness (when compared with each other within an organism) but we must not run together possession of individuality (an ingredient of which is fitness variance) with possession of fitness itself.

5.6 Fitness *variance* is exported to the higher level during an ETI

While I have raised numerous issues with Fitness Export theory above, I suggest there are two major mistakes that can be identified. One is that fitness-bearing is constitutive of biological individuality. The other is that ETIs involve a fitness self-sacrifice (on the part of particles) for the greater (collective) good.

First consider the idea that fitness-bearing is constitutive of biological individuality. While I argue against Fitness Export Theory's claim that 'during an ETI, fitness is exported to the higher level', I offer a modified version of this aphorism that is superficially similar in wording but rather different in meaning. I claim that: **during an ETI, fitness *variance* is exported to the higher level.** (This is not the whole story but it is part of it.)

If one accepts that Price's character-fitness covariance approach to evolution is appropriate, then one is obliged to accept that fitness *variance* is, by definition, partitioned differently before and after a transition in the level of individuality. (For the fitness of entities at some level to co-vary with their heritable characters, the fitnesses of those entities must vary - see Table 1.1.) This much is uncontroversial. What is controversial, I contend, is the assumption that a transition in the hierarchical level at which biological individuality obtains is implied by a transition in the hierarchical level at which the very property of fitness obtains. In my view, this

assumption (made by Michod and others) runs together the concepts of fitness and individuality, so generating the pseudo-problem of explaining how collective fitness$_2$, decoupled from particle fitness and its aggregates, comes into being as a new property as the result of an ETI. Heritable fitness variance is constitutive of individuality, fitness per-se is not[7].

I said above that a shift in the level at which fitness variance obtains is not the whole story. As discussed in Section 1.3, collective-level heredity is also needed (in addition to collective-level fitness variance) if ETIs are to yield new evolutionary units at the higher level (and not just new units of selection in the narrow sense). In terms of the Price Approach, we can say that an effective summary of what happens during an ETI would need to address both a move from particle-level to collective-level fitness variance (i.e. from $cov(w, z)$ to $cov(W, Z)$) and also an increase in collective-level heredity (i.e. that $Cov(Z, Z')$ should go from zero to something substantial).

Some commentators might consider my objections to fitness export theory to be based on mere wordplay, that the insertion of the word 'variance' is a trivial change. The long history of publications does not support this reading of Fitness Export Theory, which relies absolutely on the idea that particles lose fitness during an ETI.

I turn now to the intuition of self-sacrifice for the greater good. I do not find that particle and collective fitness[8] interests are necessarily opposed. Evolutionary transitions, I argue, involve the social niche (i.e. the game being played) being changed such that particle and collective fitness interests (where collective fitness interests are understood as aggregated particle fitness interest and not some ontologically new kind of fitness interests that comes into being in a mysterious way) are brought into alignment. This is a view of ETIs as driven by self-interest and unintended consequences rather than self-sacrifice (in common with other evolutionary change). Granted, there is a certain tension involved when we observe the evolution of a behaviour that *would have been* costly in the counterfactual case where social niche modifiers had not acted and fitness interests of particles were not aligned. However, the tension is due to our failure to take into account the fact that changing assortment changes the game, leaving pro-social behaviour individually as well as collectively beneficial. I will return to this tension in the Discussion.

If, then, fitness and individuality are different things, and individuality is understood to involve fitness-variance as a necessary ingredient, this still leaves unexplained fitness itself (and the relation between particle fitnesses and collective fitness$_2$). That is the topic of the next chapter.

[7] In the next chapter I develop a fitness concept that makes sense even in the absence of fitness variance.
[8] Collective fitness$_1$

Chapter 6

Fitness and evolutionary transitions in individuality

6.1 Overview of chapter

In the previous chapter I was critical of the view that evolutionary transitions in individuality can be explained in terms of altruistic sacrifice of fitness on the part of particles, so that the property of fitness is transferred upward to collectives of those particles. I was also critical of the view that fitness-bearing is constitutive of biological individuality. The task of that chapter was largely negative, arguing against the currently best-accepted theory - Fitness Export Theory. This chapter's project is constructive, arguing for a revised conception of fitness distinct from that employed in much theoretical work in population biology and evolutionary transitions. I suggest this revised fitness concept is an improvement because it is able to feature in a coherent explanation of evolutionary transitions that does not face the problems and paradoxes discussed in the previous chapter. I also offer an explanation for how theoretical evolutionary biology could have come this far with a problematic understanding of its central concept.

I start with a helicopter review of the role of fitness in evolutionary theory. Fitness is a vexed concept even before we consider transitory individuality but I sketch out the main roles the concept plays (as cause, descriptor and predictor) and how they relate to one another. Then I zoom in on the role fitness plays in theoretical work on ETIs - typically relative offspring count over a single non-overlapping generation. I explain why neither this fitness concept (nor any close variants) can do the work that is required of fitness in a theory of evolutionary transitions in individuality.

My main move is then to argue that something akin to the fitness concept advocated by Leigh Van Valen deserves to be resurrected for this purpose. Van Valen's fitness is about occupancy of the world through space and time. This is an absolute quantity (not a relative proportion), it does not involve counting pre-individuated 'things', it is taxon-neutral, time-scale neutral and it is

neutral with regard to hierarchical levels of the part-whole (particle-collective) kind discussed in this thesis. (Furthermore, Van Valen's unificatory fitness concept does not privilege reproduction over persistence and growth, it sees all these processes as contingent ways to achieve fitness. That means it is able to dissolve some of the puzzles surrounding problematic cases of biological individuality discussed in Chapter 2.)

While it is possible to formulate abstract models of social niche construction theory without introducing Van Valen's fitness concept, the two form a consistent package that is intuitively appealing. I will make the case that Van Valen's fitness concept (or a close variant) has a central role to play in a consistent theoretical treatment of ETIs (or, at least, that it can help us to overcome the barriers to understanding put in our way by incumbent ways of thinking about fitness).

Given that Van Valen's fitness concept is so different from the incumbent, I will need to explain how we got so far with a problematic conception of fitness, including why Fitness Export theory seems so plausible to many (including my former self). That is how I conclude the chapter.

6.2 Various fitness concepts are employed in evolutionary theory

It is conventional for discussions of what fitness actually is to begin by stating that it is a concept both central to evolutionary biology and deeply problematic (Levins, 1968; Mills and Beatty, 1979; Dawkins, 1982; Paul, 1992; Beatty, 1992; Ariew and Lewontin, 2004; Krimbas, 2004; Godfrey-Smith, 2009; Rosenberg and Bouchard, 2010). As Levins (1968) put it, "Fitness enters population biology as a vague heuristic notion, rich in metaphor but poor in precision." A full treatment is beyond the scope of this thesis. For reviews, see Sober (1984, 1993, Chapter 3), Ariew and Lewontin (2004), Godfrey-Smith (2009, pages 27-31) and Rosenberg and Bouchard (2010). In this section I offer sufficient overview to be able to show what roles fitness performs in evolutionary theory, particularly where ETIs are concerned - and what the problems are. The problems arise partly because the word 'fitness' has been used to denote several different concepts but also, I will claim, because some of the concepts are misleading or wrong.

Fitness can be (a least) a cause of, a descriptor of or a predictor of an elusive quantity known as 'reproductive success' (itself also requiring significant elaboration)[1]. Most of what I have to say is about problems with the way 'reproductive success' is understood in contemporary evolutionary theory (and Fitness Export Theory) and a proposed way forward. First, for context and to show connections with the wider literature on fitness, I will briefly survey some other potential confusions relating to fitness.

[1]While Darwin (and many others since) privileged reproduction as the main or only way to achieve evolutionary success, I intend the term 'reproductive success' here to be understood very broadly, including ways to be differentially successful other than differential reproduction in the narrow sense (this will include differential expansion and differential persistence). Where confusion is possible, I will use the term 'broad-sense reproductive success' for evolutionary success, postponing precision on what that means until Section 6.5.

First there is the distinction between fitness as cause and effect, what Sober (2001) calls the 'two faces of fitness'. When fitness is taken to be a relation between an organism and its environment, it plays the role of a cause in evolutionary explanations. Organisms can be differentially fitted to their environment and such differences can cause differences in outcomes. This concept was invoked by Darwin as both the explanandum and the explanans of his theory (Ariew and Lewontin, 2004). (See Reeve and Sherman (1993) for a review of the role of fittedness in evolutionary theory.) Paul (1992) points out that, with the arrival of the theoretical approach of Haldane, Wright and Fisher, fitness became not a *cause* of reproductive success but reproductive success itself[2]. When fitness is taken to be not a cause but an effect, it plays the role of representing something like reproductive success - an outcome.

Many different interpretations of fitness-as-outcome ('reproductive success') are possible and controversy around them continues; see Ariew and Lewontin (2004) for a review. This controversy is not directly relevant to my project here but I discuss it briefly for the sake of being clear about what I am (and am not) talking about. Abrams (2012) provides a very useful classification of fitness concepts as follows:

- **token fitness** (a property of a particular individual)
  - **measurable token fitness** (offspring count of an individual)
  - **tendential token fitness** (the propensity for an individual to have a certain amount of reproductive success in a given environment)
- **type fitness** (a property of a genotype or phenotype)
  - **statistical type fitness** (calculated from a dataset of measured token fitnesses of many individuals)
  - **parametric type fitness** (an unobservable quantity, of which statistical type fitness is an estimate)
- **purely mathematical fitness** (a quantity that features in models, including those of ETIs, and that is neutral about the distinction between the previous four categories)

Measurable token fitness is something like empirically observed offspring count for a specific individual (e.g. Dobzhansky, 1970). It is partly caused by the relation between the properties of the individual and its environment but it is also partly caused by chance events (e.g. lightning strikes), frequency-dependent effects and other confounding factors (Mills and Beatty, 1979). Statistical type fitness is worked out by performing statistical analysis on a dataset consisting of many measured token fitnesses, in the hope that random noise will cancel out and a systematic pattern will be revealed. Parametric type fitness is an underlying unobservable property of a type that statistical type fitness allows us to estimate. (This is a scientific realist concept, those who

[2]The organism-environment relation described by ecological fittedness barely features in theoretical population biology at all. See Gardner (2009) for discussion of this point and Dobzhansky (e.g. 1970, Chapter 3), Hartl and Clark (1997) and Gillespie (1998) for examples.

invoke parametric type fitness take it to exist independently of our knowledge of it). Tendential token fitness is the fitness concept at work in the much-discussed propensity interpretation of fitness (Brandon, 1978; Mills and Beatty, 1979). It is the expected reproductive success of an individual due to the relation between its type and its environment (where 'expected' and 'reproductive success' require substantial elaboration and are subject to controversy beyond the scope of this thesis - see Byerly and Michod (1991) for a criticism and Mills and Beatty (1994) for a defence of this interpretation of fitness). Each of these concepts is problematic in some way (Ariew and Lewontin, 2004; Abrams, 2009). Debate rages on whether fitness differences understood in each of the first four ways can explain evolutionary change and especially whether they can explain adaptation (for a review see Sterelny and Griffiths, 1999). Fortunately, none of what I have to say here turns on these distinctions.

What I have to say in this chapter is primarily about Abrams' fifth category, 'purely mathematical fitness'. This fitness concept appears in many theoretical models of the evolutionary process (e.g. Michod, 1999; Rice, 2004; Nowak, 2006; McElreath and Boyd, 2008). It is neutral about whether it pertains to types or tokens and whether it is a fact about the past or an expectation of the future; it can be used in all of those contexts according to the modeller's intent. However, there are several aspects of the way purely mathematical fitness is used in theoretical models that are *not* neutral with regard to what fitness is - particularly their assumptions about countability, about relative versus absolute numbers and about time scales - and all of these are important for a theory of ETIs.

6.3 Theoretical work on ETIs inherits a relative-offspring-count fitness concept from population biology

In this section I draw attention to a number of simplifying assumptions often made in theoretical evolutionary biology and carried over into theoretical work in ETIs. The survey is not intended to exhaustively cover all such models (a near-impossible task) but simply to draw attention to certain very widely used modelling assumptions that I will subsequently show to pre-judge a number of questions relevant to ETIs. I identify the fitness concept at work in theoretical work on evolutionary transitions, including Fitness Export theory, as relative offspring count - usually over a single non-overlapping generation (e.g. Michod, 1999; Okasha, 2006). This fitness notion is an example of 'purely mathematical fitness' in Abrams' above taxonomy and is derived from classic models in population genetics due to Fisher, Wright, Dobzhansky and others.

Population genetics concerns itself primarily with working out likely changes in the relative frequencies of genotypes in a population over time (e.g. Dobzhansky, 1970). The classic and influential models of selection and drift in population genetics due to Wright (1931) and Fisher (1930) typically involve a population of fixed size and non-overlapping generations. Strictly, the original Wright-Fisher model (Fisher, 1930; Wright, 1931) involved only drift, not selection and included diploidy and recombination. I use the term 'Wright-Fisher Process' in the broader

sense in which it has come to be used since, to refer to a class of models with a fixed population size, non-overlapping generations and where places in the offspring population are allocated by sampling with replacement from the parent population according to some probability distribution. That probability distribution usually depends on properties of the individuals in the population that are of interest to the modeller. This type of model is well-studied. If the variation in the population is selectively neutral, then drift occurs. If some variants are more likely than others to be selected for inclusion in the next generation, then selection occurs and the relative proportions of the types change. The dynamics of such a system, while a greatly simplified model of biological reality, can be captured in tractable mathematical models (e.g. Dobzhansky, 1937, 1970, p.101). Fitness of a focal type in such a model is thus the rate of change of the type compared with the population mean, over a specified time frame (most simply this is over a single non-overlapping generation). The dynamics are largely independent of absolute numbers or population size (Hartl and Clark, 1997, p.214). (I say 'largely' because usually population size is not itself modelled but it can influence such models insofar as it impacts on the variance of stochastic processes.)

There is an alternative type of model also commonly employed in theoretical evolutionary biology, due to Moran (1958). Models based on a Moran Process assume a fixed population size and overlapping generations. (Sometimes, quite appropriately, Moran Process models are called birth-death updating models). In each time step a single individual is chosen to reproduce, based on some probability distribution over the population. In each time step, a single individual is also chosen to die, again based on some (other) probability distribution over the population. Each of those distributions usually depends on properties of the individuals in the population that are of interest to the modeller. Moran Process models lend themselves to certain kinds of mathematical analysis more than Wright-Fisher models (Nowak, 2006). As with the Wright-Fisher model - if the genetic variation is selectively neutral, then drift occurs. If some variants are more likely than others to be selected for reproduction, then selection occurs and the relative proportions of the types change.

While the details of the Moran Process differ from the Wright-Fisher Process, the notion of fitness (i.e. broad-sense reproductive success) at work in both is a relative one. There is a fixed population size and types compete for representation in it. In a Wright-Fisher Process, differential representation arises due to differential reproduction (ignoring random effects). However, in a Moran Process, differential representation can arise in two ways. It can arise due to differential reproduction, as in the Wright-Fisher Process. It can also arise due to differential persistence (due to differential death rates) - a possibility excluded from the assumptions of the Wright-Fisher Process. In a Wright-Fisher type model, time is modelled as discrete non-overlapping generations. The state of the population in each generation is a function of its state in the previous generation plus the dynamics of the model. In a Moran Process model, time is not modelled as discrete generations - many individuals persist for multiple time steps. In both the Wright-Fisher and Moran Processes, all individuals are counted equally. That is, differential somatic growth of individuals is not modelled as way to get differential representation in the population

in either of these classes of model. (The relevance of all this is that I am shortly going to present Van Valen's view that differential reproduction, differential somatic growth and differential persistence are all contingent ways to *realize* differential fitness, which is itself something more basic.)

Models of ETIs

Both the Wright-Fisher and the Moran Process style of models are routinely used in theoretical work on the evolution of cooperation and work relevant to the major transitions. Recent examples based on the Moran Process include Lieberman et al. (2005), Ohtsuki et al. (2006), Traulsen and Nowak (2006), Fu et al. (2010) and Ji and Xian-Jia (2011). Recent examples based on the Wright-Fisher Process include models in Sober and Wilson (1998); Michod (1999); Skyrms (2004); Fletcher and Zwick (2004); Okasha (2006); McElreath and Boyd (2008); Godfrey-Smith (2008); Powers et al. (2011); Van Dyken and Wade (2012); Clarke (2014) and this thesis.

Fitness Export Theory is supported with models based on a Wright-Fisher type process (Michod et al., 2006; Michod, 2006; Okasha, 2009). The understanding of fitness at work in Fitness Export theory is one that is relative and is based on counting. For example, Michod (1999, p.175) explicitly defines his fitness concept ('F-fitness', after Fisher) as "the per capita rate of increase as it is causally determined by a genotype's heritable capacities, reproductive system, and genetic system in systematic interaction with the environment. ... In the modelling of natural selection, we seek a fitness concept that embodies the systematic components of the transformation in frequency of a type during a generation." It also involves counting offspring (each of which is equally weighted) and doing so over non-overlapping generations (Michod, 1999, 2005; Michod et al., 2006; Michod, 2006). In the next section I will explain why this model of reproductive success is unsuited to explaining ETIs.

6.4 Problems with the fitness concept in ETI theory (counting and relativity)

In this section I show that the conception of fitness (i.e. broad-sense reproductive success) used in theoretical work on ETIs is not fit for purpose. It is problematic because it is based on counting offspring - and counting things pre-supposes that those things are sufficiently individuated to be countable. Individuation is the thing to be explained, so it cannot also be used as an assumption.

The practice of performing fitness comparisons by counting things also pre-supposes that those things each bear equal quanta of fitness - and so it presupposes that particle and collective fitness are incommensurable.

The understanding of fitness as a relative quantity is also problematic if one accepts that ETIs involve an absolute increase in social welfare - an idea to be elaborated below. There are further problems with the measurement of fitness at different timescales at different hierarchical levels.

6.4.1 The assumption that fitness involves counting things is overly narrow and is question-begging

Many models of evolutionary processes used in theoretical work on ETIs take fitness (i.e. broad-sense reproductive success) to be a function of offspring count. In this section I show that equating fitness with offspring count (or using offspring count as a proxy for fitness), is multiply problematic. Some problems are due to a failure for this modelling assumption to correspond to the real-world phenomenon it models in the majority of cases[3]. Offspring counting does not sit comfortably with entities having problematic status as biological individuals due to a blurring of reproduction and somatic growth, nor with entities that differ in fitness primarily due to differential persistence. Another category of problem involves flawed explanatory logic. Importantly for work on ETIs, offspring counting is problematic because the very act of counting things presupposes there are things to count, so cannot apply to collective-level entities at the beginning of the evolution of their individuality. Furthermore, the practice of counting things for the purpose of comparison presupposes that counts of those things can meaningfully be compared. For that criterion to be met, they need to be things of the same type and equivalent value. (The idiom 'comparing apples with apples' captures this idea, especially if we add a qualification that we are talking about a standardized apple). Such a practice therefore precludes any commensurability of fitnesses at different hierarchical levels (e.g. particles and collectives) or between species (e.g. inter-specific symbioses such as lichens). This is a problem for evolutionary theories of fraternal or egalitarian transitions in individuality, respectively.

In a Wright-Fisher type model with constant population size, non-overlapping generations and subsequent generations constructed by sampling individuals with replacement from the current generation, there is only one way for types to be differentially fit. That is for a type in generation $t + 1$ to make up a different proportion of the population than it made up in generation t. (Another way to say the same thing is that types can be differentially fit by having a per-capita rate of increase that differs from the population's mean per-capita rate of increase.) All offspring are counted equally and all live for the same time (a single generation). This simple idealized model captures some important aspects of evolutionary dynamics but, like all models, it is a special case that does not model all biological possibilities. This sort of model is applicable to countable organisms in a niche that supports a constant population size (perhaps because it is at a stable carrying-capacity) and that have synchronised generations and that reproduce once then die (i.e. semelparous univoltine organisms such as mayflies and others in the order ephemeroptera). However, this is a fairly narrow class of organisms. As we saw in the literature review (Chapter 2), counting individuals (and therefore offspring) is problematic in the many taxa where individuality is not well defined. This category (previously referred to as non-Weismannian organisms) includes many animals (outside the chordata), plants, fungi and colonial organisms having problematic status as biological individuals. Far from being an

[3]Recall Buss's complaint, discussed in Chapter 2, about the inappropriate use of the concept of the 'Weismannian organism'.

exotic curiosity, most known species are in this category[4]. The assumption that counting must be involved in fitness measurement therefore runs up against a familiar problem of biological individuality - that of distinguishing narrow-sense reproduction from somatic or clonal colony growth (Harper, 1977; Janzen, 1977; Jackson and Coates, 1986; Clarke, 2010).

Why does counting matter?

Rejecting the role of counting is important because counting assumes that the things being counted are all equally weighted, which precludes commensurability of fitnesses at different hierarchical levels (such as particles and collectives). Therefore, the assumption that counting offspring can be used as a proxy for fitness pre-judges the question of whether particle and collective fitnesses are decoupled. It pre-judges the distinctness of collective fitness$_1$ (aggregated particle fitness) and collective fitness$_2$ (collective fitness in its own right) that is at the conceptual heart of Fitness Export Theory.

There are further problems. Using offspring count as a proxy for fitness precludes commensurability of fitness across species boundaries and precludes measurement of the fitness of organisms involving inter-specific symbioses, which is most of them (Dupré, 2010; Bouchard, 2013).

The assumption of incommensurability between hierarchical levels pre-judges questions relevant to the fraternal evolutionary transitions, while the assumption of incommensurability between species pre-judges questions relevant to the egalitarian evolutionary transitions. As Bourrat (2015a,b) has pointed out, the assumption that particle and collective fitnesses can be measured (by counting) on different time scales (e.g. Okasha, 2006, p.58) pre-supposes the distinction that is supposed to be being explained in Fitness Export Theory.

Counting and timescales

Modeling fitness as offspring count can fail to correspond to the world in other ways. It is not normally the case that generations are synchronized. Even with equal numbers of offspring, a type reproducing more quickly with respect to time will increase in frequency relative to another (Godfrey-Smith, 2009, Chapter 2). In age-structured populations where there is cummulative exogenous mortality risk, a reproductive bird in the hand is worth two in the bush (Charlesworth, 1980). There are also more subtle ways the idealized offspring-count conception of fitness can fail to correspond to actual-world fitness. For example, even if generations are synchronized, it is sometimes better in the long run to produce less offspring with a low variance in offspring count than to produce more offspring with a higher variance, especially when the effective population size is small (Gillespie, 1974, 1977). In such cases, the timescale over which fitness is measured matters, as does the effective population size.

Differential fitness can arise through differential persistence of the same individuals through time, rather than differential reproduction of new individuals. For example, many plants (Fox,

[4]Buss' (1983) Table 1 enumerates the main taxa in the tree of life (as it was understood at the time) and shows that the received view of the biological individual - Weismannian organisms - applies only to a tiny portion of it, mainly in the chordata. These few taxa are amenable to an offspring-counting analysis of fitness differences between individuals. The rest are not.

1990) and mycelial fungi (Burnett, 2003) employ perennating structures - typically underground rhizomes or tubers (plants) or dense bundles of hyphae (fungi) from which the organism can recreate itself in the event that the more expansive parts are destroyed by unfavourable conditions. Perennating structures are effectively propagules that extend the organism's existence through time just as seeds and spores extend the organism's kind's existence through space.

The strategy of model-based science

Godfrey-Smith's (2006) discussion of the 'strategy of model-based science' (first mentioned in the previous chapter in a different context) provides a useful framework for thinking about what the relative-offspring-count conception of fitness gets right and what it does not. On Godfrey-Smith's account, the strategy of model-based science is to construct a model known not to correspond to the actual world but 1) having a causally transparent mechanism and 2) bearing a relation of relevant similarity to a particular aspect of the actual world in such a way that some understanding of that aspect of the actual world is conveyed by understanding the model. Simple models like those based on the Wright-Fisher Process or Moran Process are causally transparent mechanisms. They are known not to correspond exactly to the actual world but, the idea goes, they convey understanding because they bear a similarity relation to the actual world in certain relevant respects. I claim that, while such models may be useful for many purposes, the similarity relation does not hold for the way fitness is modelled and the way it is in the world, especially in the context of evolutionary transitions in individuality.

Counting pre-supposes individuation, which is the thing to be explained by ETI theory

Theoretical work on ETIs has the goal of explaining, in Darwinian terms, how new levels of Darwinian individuality can come into being in the world as a result of a Darwinian process acting at a lower hierarchical level. A theoretical framework that pre-supposes individuation is therefore problematic. Any fitness notion based on counting individuals is bound to run into trouble when individuality is in transition. The key to that problem, I suggest, is to recognise that the basic notions of offspring and counting are themselves problematic in this context. The very practice of offspring-counting assumes that the things being counted are countable and that reproduction (and with it the parent-offspring relation) is basic. It also assumes that fitness is had via reproduction alone (or that it can be represented as such without loss of generality). These assumptions are not generally true but many theoretical approaches to ETIs, including Fitness Export Theory, assume them. Indeed, even among ETI researchers, it is widely held that counting is central to fitness (Clarke, 2011a,b).

6.4.2 Changes in relative proportions fail to model changes in absolute numbers

A central idea in the theory of major evolutionary transitions is that **collective action is *not* a zero-sum game** but that betterment is possible (subject to overcoming certain problems connected with free-riding and so on, discussed at length in Chapter 3). Models that assume a fixed population size implement a zero-sum game between types for inclusion in the population. Such

an approach considers only changes in the relative proportions of types in a population and not gross changes in the size of that population, so fails to apprehend an important part of the thing to be explained. (Social Niche Construction theory has it that, ceteris paribus, absolute fitness can be raised by evolutionary transitions and that is often what explains their occurrence. I return to this theme below in Section 6.6.)

For example, the evolution of multicellularity in the Charophyte algae paved the way for the colonisation of the land by plants (Umen, 2014). Let us suppose that the absolute number of cells in the clade was greatly increased by this innovation. An analysis of the relative frequencies of certain alleles in the global population of Charophytes and their descendants would be missing something important if it failed also to take account of the significant change in the population size. (For further examples, see Platt and Bever (2009) and West et al. (2007a) who provide examples from the evolution of cooperation in microbes, involving the production of public goods that enable population expansion).

6.5 Van Valen on fitness as occupancy of the world through space and time

Van Valen (1976; 1980; 1989; 2003) argued for a novel theory of fitness that solves many of the problems with thinking of fitness as relative-offspring-count, enumerated in the last section[5]. It is not relative, it does not consider fitness to arise only through the begetting of offspring and it does not rely on counting things.

> *"Every living thing is a sort of imperialist, seeking to transform as much as possible of its environment into itself and its seed."* - Russell (1927), quoted by Van Valen (1976)

The above quotation from Russell captures the intuition behind Van Valen's theory of fitness, which he sees as success in colonisation of the world through both space and time, howsoever this is achieved. This is a single, unifying conception of 'evolutionary success' that applies commensurably to all life in all taxa and at all hierarchical levels.

> "I . . . have defined the realized absolute fitness of any evolutionary unit roughly as its control of trophic energy. . . . [T]he critical point here is the replacement of number of individuals with energy control. . . . The use of energy makes all forms of natural selection commensurable and permits unified analysis at all levels." Van Valen (1976)

[5] I am indebted to Carl Simpson (personal communication) for introducing me to the ideas of Van Valen in the context of ETIs.

> "Fitness is the central concept of the evolutionary half of biology, yet it is an exceedingly elusive concept. In population genetics it is the number of offspring, usually one generation later, and this notion is usually adequate within the domain of genetics. But when the domain is expanded the inadequacy of considering only numbers of individuals becomes apparent. Individuals differ in their size and other aspects of quality, and even the boundaries of an individual are arbitrary in many kinds of organisms. The means of control of the numbers of individuals in a population is ignored, yet this can strongly interact with their quality. And so on. In any real case natural selection has a single and well-defined outcome, so the problem is in our analysis rather than in the real world. Energy provides the unifying mechanism. In the usual domain of population genetics it reduces to the standard concept as it should, but it permits a single analysis at all levels. It permits such a single analysis because it drives all processes and is the single ultimate currency." Van Valen (1980, p.292)

For Van Valen, evolutionary success is ultimately about control of energy for further expansion through space (via reproduction and somatic growth) and through time (via persistence). Energy-control is thus the most basic fitness concept. It is an absolute, conserved, non-countable physical quantity, measured in joules. Competition for non-energetic resources (e.g. potassium and phosphorus in the rhizosphere or space on a rocky shore) exists because those resources are required as a means to acquire control of energy but are not ends in themselves.

Differential reproduction (differential offspring-count), differential somatic growth (differential expansion) and differential persistence are all contingent ways to instantiate differential Van Valen fitness.

At various times other authors have entertained views partially consistent with or borrowed from Van Valen. Ariew and Lewontin (2004, Section 6) briefly sketch out a theory of fitness as 'occupancy of the external world' that is similar to Van Valen's[6]. Bouchard (2008, 2013) and Bourrat (2015a) argue for differential persistence as a realiser of differential fitness, while Simpson (2011a,b) argues for differential expansion as a realiser of differential fitness. For an explicitly gene-oriented view of fitness following Van Valen, see Williams (1992, page 18). See Rosenberg and McShea (2007, pp. 140-141) for a discussion of Van Valen's theory of fitness. They conclude that "[this] viewpoint leads to certain conclusions that are highly intuitive and yet run contrary to the standard view of fitness, especially in the evolution of groups and in evolution on long time scales.".

[6] Ariew (personal communication) credits Van Valen as the originator of the idea.

Caveats

1) Some commentators on Van Valen have taken his fitness concept to pertain to mass rather than energy (e.g. Williams, 1992; Clarke, 2011b) or, more vaguely, to occupancy of the world (Ariew and Lewontin, 2004, Section 6). Perhaps control of energy is a propensity while realised occupancy of the world is an outcome. While I find Van Valen's views on the single unifying basis of fitness persuasive, what I have to say in this thesis does not depend on identifying fitness with any particular physical quantity. I am interested in Van Valen fitness for its unifying role, due to its being absolute and commensurable across all life at all spatial and temporal scales.

2) For Van Valen, the beneficiaries of fitness are, at bottom, genes:

> "... consider all the [gene instances] in a community, all the allele copies. Some of these genes, or their replicas, will be there at some later time - others won't - and frequencies will have changed. This gives a view of absolute fitnesses at the genic level, and species [and individual, and collective] boundaries are irrelevant here even though partly relevant in causing the differences. ... Some of these genes will have remained because of persistence of single long-lived individuals, some will have increased because an individual has grown, and some will have increased by reproduction or dispersal." Van Valen (1989)

It is not important for what I have to say here that genes are the ultimate bearers of fitness - I just need there to be some basic level where all fitnesses are commensurable in absolute terms. Van Valen intends his fitness concept to be fully consistent with inclusive fitness theory. He differs from mainstream biologists in what he understands fitness to be, rather than who may be said to have it when entities share genes. I do not enter into the controveries about the gene's eye view, book-keeping argument, etc. (Sterelny and Griffiths, 1999). For my purposes, the important message is that fitness is something more basic than particle fitness and that that basic thing does not change in its ontological character if its variance becomes differently partitioned due to an ETI.

I will use the term 'VV-fitness' to refer to a fitness concept of this type: a single conserved physical quantity that is absolute in value and commensurable across all life. It can be understood as applying to types or tokens, as historic outcome or propensity, or as a purely mathematical concept used in modelling (Section 6.2). The term is not intended to commit the user to identifying VV-fitness with a particular physical quantity (such as energy or mass), nor is it intended to commit the user to an explicitly gene-centred interpretation. This is not to deny either of those connections but merely to declare debate on them to be outside the scope of this thesis.

6.5.1 Differential VV-fitness can be due to differential reproduction (differential offspring count)

In idealised cases of the kind modelled by the Wright-Fisher process (Section 6.3) where population size is fixed, individuation is clear (Weismannian), generations do not overlap and types compete for representation in the next generation, VV-fitness for a type (or token of a type, depending on what is being modelled) can be maximised by maximising its Fisherian fitness (i.e. per-capita rate of increase), as is familiar. This is what Van Valen (1989) means when he says "In the usual domain of population genetics it reduces to the standard concept, as it should". However, as I showed in the previous section, the simple models of fitness used in theoretical work on ETIs (involving counting and relative proportions) are unsuitable for a theory of evolutionary transitions in individuality. I suggest that Van Valen fitness is a plausible way to understand evolutionary success (i.e. broad-sense reproductive success) that does not suffer from those flaws.

On this view, counting offspring can be an effective way to compare the fitnesses of like-kinds (assuming they are countable and that each one is of roughly the same reproductive value and that reproduction by offspring is their only way of getting fitness - a circumstance approximated only by Weismannian organisms.) Importantly, counting is now a shortcut, a heuristic device that can be useful when the circumstances allow it (e.g. with mayflies) but not generally so.

Differential relative-offspring-count is sometimes a *realizer* of differential VV-fitness. But to assume relative-offspring-count *is* fitness is to mistake the contingent for the essential.

(Counting offspring is still rather more practical for the empirical biologist than capturing them all and burning them in a bomb-calorimeter to estimate the quantity of energy they contain.)

6.5.2 Differential VV-fitness can be due to differential expansion

Van Valen (1989) reports that he discovered the following quotation (mis-attributed to himself) written on a toilet wall at the University of Chicago: "fatness is fitness". Outside the Weismannian organisms, there is some truth in this. Other things equal, Van Valen has it that a number of larger organisms, controlling more expansive energy, is fitter than the same number of smaller organisms, controlling less. This is true whether the organisms are conspecifics or not. Speaking to the problems connected to assuming an arbitrary distinction between reproduction and somatic growth discussed in Section 6.4.1 above, Van Valen (1989, page 10) says "...in many somatogens [i.e. non-Weismannian organisms, without segregated germlines] there is no sharp distinction between growth and reproduction; think of a tillering grass or a budding bryozoan or a clone of cladocerans [water fleas with a mainly asexual lifecycle similar to that of aphids]. It is therefore dangerous to use a basic theoretical framework which depends on such a sharp distinction and which must create one arbitrarily even when it doesn't exist."

Van Valen provides the following important example of how a focus on narrow-sense reproduction (offspring-counting) at the expense of expansion can fail adequately to model evolutionary change (Van Valen and Sloan, 1966; Van Valen, 1989). Species selection is conventionally understood as a process driven by variation in number of offspring species (i.e. collective fitness$_2$) covarying with species character. From the fossil record, we known the multituberculates were a clade of rodent-like mammals that lived in the Mesozoic epoch and that went extinct about 35 million years ago. They were out-competed by a newer type of mammal, the placentals. During the decline of the multituberculates the clade diversified (speciated) many times, yielding an ever larger number of multituberculate species. Yet all the while, the absolute number of multituberculates (and - more importantly - the absolute amount of expansive energy they controlled) was going down, while the placentals gained control of more and more expansive energy. An understanding of fitness based on offspring count fails to account for such a situation, while an account based on VV-fitness accounts for it straightforwardly[7]. As discussed in Section 6.4.1 above, the assumption behind counting is that each thing being counted has the same value and this is not generally true.

As some authors have noted, expansion at one hierarchical level amounts to extra reproduction of the units at the level below (Clarke, 2011b; Simpson, 2011b). This reflects the blurred distinction between reproduction and growth in nature. Many of the conceptual problems surrounding biological individuality of clonal and colonial organisms (e.g. the ramet/genet counting problem in botany or similar problems with clonal marine invertebrates discussed in Chapter 2) dissolve once a Van Valen-like fitness concept is adopted.

There are many examples of collective lifeforms where collective fitness is proportional to particle fitness. For example, Mendoza and Franco (1998) studied a tropical palm (Reinhardtia gracilis) with a two-level population structure that could be modelled as particles nested inside collectives (Figure 1.1). The palms grow in clumps (collectives) on the forest floor. Each clump is a single genet - a clonally formed group of ramets (particles). Mendoza & Franco found that: “Ramet production increases genet size and this in turn increases genet reproductive performance. Clonal growth in this species may be viewed as a growth strategy that tends to maximize genet fitness.”

The encrusting cheilostome bryozoans are a taxon of colonial marine invertebrates which form a crusty coating on rocks and other solid surfaces on the sea bed in tropical waters. They have a two-level population structure in which the zooids play the role of particles and the colonies the role of collectives. Winston and Jackson (1984) report a correlation between colony size (i.e. number of zooids) and number of larvae generated (colony reproductive output).

Nakaya et al. (2003) provide another example, this time from work on colonial ascidians (sea squirts). Sea squirt colonies can be viewed as collectives while particular sea squirts can be viewed as particles. The colony’s reproductive output (capacity to found new colonies) is an increasing function of the number of sea squirts in it.

[7] See Sober (1984, pp. 367-368) for a similar example involving grasshoppers.

6.5.3 Differential VV-fitness can be due to differential persistence

> "It is just as good, and maybe better, for a massive coral or a tree to stay alive, occupying the same good site, as it is for it to reproduce into an uncertain world." Van Valen (1989)

Differential occupancy of the world can be a result of differential persistence of entities even when there is no difference in their reproductive output measured as number of offspring. To see how, consider a patch of land in a climactic-cover forest that has been vacated by the death of an old tree. In case 1, a seedling of genome Y takes hold and grows to maturity; we come back after 500 years and find that Y is still alive and vigorous. In case 2, a seedling of genome Y takes hold, grows to maturity, leaves one offspring in the same place and dies; we come back after 500 years and find that descendent of Y alive and vigorous. On an offspring counting understanding of fitness, we would say the fitness of Y in case 1 is zero and that it is one in case 2. However, the realpolitik is that in each case the imperial ambitions of Y have been achieved to roughly equal degrees. (Indeed, if there are any fitness differences between the two cases, it would be more reasonable to say that Y enjoys greater inclusive fitness in case 1, due to sexual outcrossing.) The message is that counting reproductive offspring is not the only way to measure fitness because reproduction is not the only way to achieve it - just staying alive is also a way to be fit in its own right (Bouchard, 2013).

Reproduction and persistence both achieve VV-fitness through different means. An emphasis on persistence over reproduction is associated with K-selected species such as elephants, while the converse is associated with r-selected species like mice (Ridley, 2003). This view of fitness also handles the case of single-celled organisms. Greater fitness can be achieved through living longer (as with m. tuberculosis) or by having a shorter g-time (as with e. coli).

In Wright-Fisher type models of evolution, the assumption of non-overlapping generations means that all entities persist for an equal time. This type of model cannot therefore model differences in evolutionary success due to differential persistence and an over-emphasis on it may lead theorists to differential neglect persistence as a way to instantiate differential fitness. Moran Process (Section 6.3) models differ in that entities persist for differing numbers of timesteps and can be differentially successful as a result of this differential persistence[8].

6.5.4 VV-fitness is commensurable across hierarchical levels and species boundaries

Van Valen is monist about fitness. VV-fitness is a single physical quantity, absolute and conserved, commensurable among all living things, regardless of their place in any part-whole hierarchies and regardless of their species (Van Valen, 1980). There is no requirement for VV-fitness

[8]Thanks to Guy Jacobs (personal communication) for pointing this out in connection with Van Valen's views on fitness.

comparisons to be made only between like kinds. Fitness comparisons between cells and multicellular organisms (particles and collectives) become possible, as do fitness comparisons that cross species boundaries. The latter point is important in the study of inter-specific symbioses and egalitarian transitions, while the former is important for the study of fraternal transitions.

In the next section I will consider the implications of this for the distinction made by multi-level selection theory (Damuth and Heisler, 1988; Okasha, 2006) between particle fitness, collective fitness as aggregate particle fitness (collective fitness$_1$) and collective fitness in its own right (collective fitness$_2$). This distinction is important here because it is key to understanding the question Fitness Export Theory takes itself to be answering.

6.6 Social Niche Construction Theory and Fitness Export Theory under VV-fitness

The fitness at work in my application of social niche construction to ETIs is something like VV-fitness. I make this connection because my account of ETIs (and my rejection of Fitness Export Theory) requires an understanding of fitness that is commensurable between hierarchical levels (I reject fitness decoupling) and that is an absolute quantity, the total amount of which can potentially be raised by cooperation (I reject the intuition of self-sacrifice for the greater good). Recall that social niche construction (SNC) theory has it that organisms are both the subject and object of their own social evolution (Chapter 3). To the extent that there are potential gains to be made from collective action, SNC theory says that changes made by entities to their own social niche that enable such collective action will be favoured by selection, if those entities themselves will preferentially benefit from any cooperation so engendered. The causal story is stated at particle level and involves natural selection favouring changes in particle characters that further the inclusive fitness interests of particles[9]. According to SNC theory, changes in social niche modifying traits (SNMs) that bring about a social niche in which more cooperative behaviour is stable will preferentially raise the VV-fitness of the bearers of those modified traits and cause them to increase in frequency (and absolutely). The VV-fitness of a collective is identical with the VV-fitness of its member particles, so someone interested in the fitness of collectives will see that collectives containing more cooperators will be fitter than those containing less. That does not necessarily mean there will be more offspring collectives (but it might - and it will for Weismannian organisms). Collectives containing more cooperators than average might persist longer, or expand to a larger size, or give rise to more offspring collectives than average (and they might do a combination of those things). That is, collectives containing more cooperators than average they will have more VV-fitness than average. This will both increase their relative representation in the population *and* increase their absolute occupancy of the world.

[9]There is no appeal to the idea of altruistic sacrifice for the greater good in SNC theory, only self-interest. That said, collective action is not a zero sum game, so the pursuit of self-interest can still raise social welfare in some circumstances - I return to this issue in the Discussion.

Stage 1	Collective fitness is identical with summed particle VV-fitness; particles are units of evolution, collectives are not (cooperation spreads among particles because cooperation is adaptive in the social niche)
Stage 2	Collective fitness is still identical with summed particle VV-fitness (and is greater than it was in Stage 1 due to benefits derived from collective action); entity VV-fitness covaries with offspring character for both particles and collectives (collectives start to emerge as evolutionary units in their own right)
Stage 3	Collective fitness is still identical with summed particle VV-fitness (and is greater than it was in Stage 2 due to benefits derived from greater collective action); collective VV-fitness covaries with mean offspring-collective character - not so for particles (collectives have fully emerged as evolutionary units)
Stage 4	Collective fitness is still identical with summed particle VV-fitness (and is greater than it was in Stage 3 due to benefits derived from even greater collective action); collective VV-fitness covaries with mean offspring-collective character; collectives evolve complex adaptations including division of labour, specialisation of parts and endless forms most beautiful

TABLE 6.1: Relation between collective fitness and particle fitness during an ETI according to social niche construction theory. Compare with Table 5.1 on page 98, which is itself based on Okasha's (2006) Table 8.1 'Relation between collective fitness and particle fitness during an evolutionary transition' above.

Recall the position I am arguing against. Fitness Export Theory says ETIs involve a move from MLS1 to MLS2, with MLS2 and fitness_2 being the defining characteristic of true collective-level individuality[10]. MLS2 features collective fitness_2, an ontologically new type of fitness, decoupled from fitness_1. The origin of collective fitness_2 stands in need of explanation. 'Fitness export' by way of altruism is then invoked to explain the provenience of collective fitness_2.

If one recognises that the use of collective fitness_2 only works for Weismannian organisms and in those cases is simply a convenient shorthand (a sort of data compression) for collective fitness_1, then things become considerable simpler and less mysterious. Rather than define an ETI as something that ends with MLS2 and fitness_2, I prefer to focus on the covariance between entity fitness and offspring character at two hierarchical levels (Section 1.4). When covariance between entity fitness and offspring character shifts from particle-level to collective-level, that is an ETI.

[10] As I explained in Section 1.3, the difference between MLS1 and MLS2 is essentially a matter of the absence of a collective-level parent-offspring relation with heredity in MLS1 and presence of it in MLS2. The difference is therefore the difference between a selection regime featuring units of selection (MLS1: $cov(w, z) \neq 0$) and units of evolution (MLS2: $cov(w, z') \neq 0$).

This approach stands in contrast with that employed in Fitness Export Theory. In my view, the need to explain the source of a new type of collective fitness is a pseudo-problem caused by adopting an inappropriate model of fitness and by running together the bearing of fitness and the possession of Darwinian individuality (of which fitness *variance* is a necessary ingredient). Loss of individuality is not loss of fitness.

I now need to explain how population biology could have got this far without noticing that fitness is not relative offspring count

6.6.1 How did population biology get this far with a faulty fitness concept?

One might find it implausible that evolutionary theory could have progressed thus far if it were based on an erroneous conception of fitness, one of its core concepts. Within the domain of population genetics, usually concerned with the evolutionary trajectory of certain traits in pre-extant populations, it is often instrumentally valid to use relative-offspring-count as a fitness proxy. This is because evolutionists really care about fitness *variance*, not fitness itself. The driver of evolutionary change is covariance of entity fitness and offspring character. This covariance is unchanged by a scaling of the fitness quantity everywhere by a constant. The ratio between VV-fitness and offspring-count is roughly constant for Weismannian organisms, so evolutionists can instrumentally 'get away with' using offspring-count-variance and get the same answer as they would have done if they had used VV-fitness-variance. For evolutionists, the important thing about fitness is that it is 'that quantity in which heritable variance is the driver of evolutionary change' and not its essential nature.

I argue that collective fitness$_2$ is simply a shortcut for collective fitness$_1$ when the collectives are Weismannian organisms (in such cases, collective fitness$_2$ is proportional to collective fitness$_1$ is proportional to VV-fitness). When all collectives are not worth the same (as with Van Valen's important example from the natural history of the multituberculates in Section 6.5.2, page 123), then collective fitness$_2$ is not the right thing to track to understand evolutionary change.

To see why fitness$_2$ can be used unproblematically by evolutionists interested in Weismannian organisms, consider the following example. There is a two-level hierarchy, particles nested in collectives. We think, based on our understanding of the natural history of the situation, that MLS2 is in operation[11]. We first measure the covariance of a collective character and collective fitness$_2$ (i.e. number of offspring collectives). We find the character to covary with fitness, with the mean change in the character over a generation being given by the Price Equation as follows:

$$\Delta\overline{z} = \frac{Cov[w, z]}{\overline{w}}$$

Secondly, we perform the same analysis, this time somehow using the count of cells (particles) in all the offspring collectives as our fitness value, i.e collective fitness$_1$. Suppose each collective

[11] Recall from Section 2.4 that the '2' in MLS2 does not imply fitness$_2$ is the only appropriate fitness measure.

is roughly equivalent in size and has approximately K cells.

$$\Delta\overline{z} = \frac{Cov[(Kw), z]}{K\overline{w}} = \frac{Cov[w, z]}{\overline{w}}$$

That is, our analysis finds the same character-fitness covariance whether we use collective fitness$_1$ or collective fitness (fitness$_2$) because, for Weismannian organisms, those things are proportional to one another. Thus, for Weismannian organisms (only) it is possible to use collective fitness$_2$ as a (very convenient and pragmatically superior) substitute for collective fitness$_1$ and get the same results. And of course collective fitness$_1$ is itself proportional to VV-fitness, which is the bottom layer. In cases where Weismann's assumptions do not hold (i.e. most life outside the chordata) then only the lower-level fitness$_1$ will be adequate (and even then this assumes each particle has equivalent VV-fitness).

This explains how it could have been possible for evolutionists to be using a special-case fitness concept without noticing that something was amiss.

Chapter 7

Discussion

In this thesis I have offered a new type of explanation for how an evolutionary transition in individuality might get started and I have offered some reasons why this explanation is an improvement on incumbent theory.

This final chapter returns to comment further on some issues the arose during the thesis. I discuss some reasons interactionist explanations (like social niche construction) may represent progress when compared to the externalist explanations usually offered in social evolution theory and I return to discuss the intuition that the evolution of cooperation involves self-sacrifice on the part of cooperating particles.

I start this final chapter with a review of the claims of the thesis (first stated in Chapter 1, Section 1.2), showing how they are supported by the intervening chapters.

7.1 Support for the claims of the thesis

The first claim listed in Section 1.2 was that some evolutionary transitions in individuality (ETIs) can be explained by social niche construction. To avoid slipperiness about what constitutes an ETI, I set up a model of what an ETI is in Section 1.4, based on character-fitness covariance shifting from one level to another in a two-level part-whole hierarchy. That enabled me to state the question like this: How does it happen that an evolutionary process involving wholly particle-level character-fitness covariance could give rise to an evolutionary process involving wholly (or mainly) collective-level character-fitness covariance? It is tempting to think about structural features of collective-level entities (such as internal conflict suppression mechanisms) as adaptations that have made those collectives into units of evolution. I wanted to avoid that route as it seems to imply that the future can cause the present - that things that don't yet exist as units of evolution could evolve themselves into existence as units of evolution. In Chapter 3, I offered an account of how social niche construction theory could meet this explanatory challenge while avoiding those sorts of metaphysical worries. SNC theory makes two key moves.

The first is to stop thinking about social dilemmas as entirely externally defined, something to which organisms must adapt themselves but which they have no power over. In the spirit of Lewontin and the niche construction theorists (Odling-Smee, Laland and Feldman), I move to an interactionist way of thinking about social evolution. A second key move is to stop thinking about evolutionary transition as an adaptation, as something selected-for and to start thinking about it as something incidental to the pursuit of fitness on the part of the entities that are unproblematically units of evolution at the start of the story (i.e. the particles). This is all done in Chapter 3, with the crux in Section 3.9. To show it's not all hand-waving, Chapter 4 describes a simulation model for the evolution of a life-history bottleneck by social niche construction.

The second claim listed is that social niche construction's interactionist explanatory stance matches the causal structure of the world better than the externalist explanatory stance, when the thing to be explained is evolutionary transitions in individuality. That claim is unpacked in Chapter 3 in the discussion of how interactionism differs from externalism and elaborated upon later in this chapter (Section 7.2). In some ways, interactionism matches the causal structure of the world better just because it is messier and the idealised causal simplicity employed by externalist explanations is rarely to be found in biology. But this is not a gratuitous messiness - it is a messiness that is necessary to the explanation, because the click-clacking between changes in social niche modifiers and changes in social traits are able to explain each other in a manner that is not available to externalist explanations. Furthermore, I take it that things that are impossible don't match the causal structure of the world, whatever it might be.

The third claim is that social niche construction provides an evolutionary explanation that proceeds without assuming the prior existence of any higher-level evolutionary process, even though the product of such transitions can be a higher-level evolutionary process. That is another way of bringing out the benefit of setting up evolutionary transitions as something incidental to the pursuit of particles' fitness interests and not something selected-for. This claim is made mainly to highlight the fact that alternative explanations actually *do* invoke higher-level evolutionary process to explain the advent of higher-level evolutionary processes - a point elaborated in Chapter 5.

The fourth claim is that the onset of an evolutionary transition in individuality is characterised by increasing alignment of fitness interests among pre-existing lower-level units, such that pre-existing lower-level units gain inclusive fitness from the transition (other things equal). Again, this might seem obvious but it has been common to claim something else - that the lower-level units actually lose fitness as part of an ETI. That issue is dealt with in Chapters 5 and 6. In Chapter 5 I critique Fitness Export theory, which claims particles necessarily lose fitness as part of an ETI. That chapter also fulfils my fifth claim, that Fitness Export theory is unworkable and does not explain what it purports to explain. The issue of altruistic self-sacrifice is further discussed in Section 7.4 below. In Chapter 6 I make the case that particles, in fact, gain fitness as part of an ETI (and that's why ETIs happen, when they do). This claim turns on how one interprets fitness in evolutionary theory. I have argued for a common-sense understanding of fitness as a measure of occupancy of the world through space and time (my sixth claim, unpacked

in Chapter 6). That means I move away from seeing fitness in an evolution-centric way (as 'that quantity heritable variation in which is the dynamo of evolutionary change') and toward seeing fitness as something that has a meaning independent of evolution.

7.2 Externalist attempts to explain ETIs are problematic

Social niche construction theory says that both social behaviour (like cooperation) and some factors modifying the social niche in which social behaviour occurs can be explained as reciprocal responses to one another (Chapter 3). That means social niche construction employs an interactionist explanatory stance. As such, it contrasts with most work in social evolution theory, where an externalist explanatory stance predominates. Recall that externalism in evolutionary biology has it that the properties of biological entities (including social behaviours such as cooperation) are to be explained in terms of their adaptation by natural selection to factors external to them. In this section I assess how well an externalist explanatory stance might handle the challenge to explain how collectives become units of evolution (and how particles lose this property) during evolutionary transitions. An externalist explanation might cast either collectives or particles in the role of evolutionary units (and adaptation-bearers), so I consider both options.

There is no doubt that many collective life-forms have striking features that suppress internal conflict (Table 3.1 on page 38 above lists many examples). Some have tried to meet the challenge by invoking an externalist explanation that casts collectives in the role of units of evolution (i.e. 'collective-level selection', properly understood). Roughly, the story is that collectives are differentially fit according to the extent to which they are successful in within-collective conflict suppression. Less conflicted collectives are fitter and so the conflict suppression mechanisms evolve as adaptations of collectives. This sort of reasoning[1] can be found in works of natural history that try to explain the existence of conflict-suppression mechanisms (Wilson and Sober, 1989; Wilson and Hölldobler, 2005; Wilson and Wilson, 2007) and in modelling work that presupposes a collective-level evolutionary process and then finds that it acts to strengthen conflict-suppression (Michod and Roze, 1999; Michod, 2006)[2]. However, I find it problematic because it invokes a product of a collective-level evolutionary process to explain how the collective level became a level at which the evolutionary process operates. This sort of reasoning seems to contravene **Williams' Principle** (Williams, 1992; Sober and Wilson, 1998; Okasha, 2006, pages 113 and 225), which states that collective-level adaptations are evidence of past collective-level selection and not a prior condition of it. As Clarke (2014) says, "Complex, late stage organismal traits may well be products of higher-level selection, but are not pre-requisites for selection

[1] For a very general study of explanations of the form 'X occurred because it would have had function Y if it did', see Cohen (1978, Chapter IX).

[2] Recall that Fitness Export Theory employs a narrative like this: the evolution of altruism happens due to Type 1 group selection, that causes fitness to be exported to the higher level, that means individuality is exported, then higher-level selection evolves conflict suppression mechanisms that legitimize the new unit once and for all. But how did the altruism evolve in the first place? This can only be because between-group selection for altruism is stronger than within-group selection against it. And the origin of between-group selection was the thing to be explained all along.

at the higher level. Simple early stage traits such as stickiness, on the other hand, are pre-requisites of higher-level selection, but they need not be products of it." Any attempt to explain ETIs that invokes collective-level adaptation[3] seems susceptible to worries of this sort, so the prospects for explaining ETIs using an externalist explanation couched in terms of a collective-level evolutionary process are not good.

Perhaps we can instead explain evolutionary transitions by taking an externalist explanatory stance invoking adaptation at particle level instead? It is hard to see how such a project could proceed. The externalist set-up treats the social niche as exogenous to the explanation. Ex hypothesi, selection favours selfishness at particle level - the opposite of what we need to explain. As Maynard-Smith (1988) asked, "How did natural selection bring about the transition from one stage to another, since, at each transition, selection for 'selfishness' between entities at the lower level would tent to counteract the change?" Any externalist attempt to explain ETIs that relies on particle-level adaptation seems to befall this sort of problem.

Social niche construction theory does not suffer from either of the two problems above. It does not invoke a collective-level evolutionary process to explain how collectives became units of evolution. Nor does it assume that a particle-level evolutionary process must always be opposed to the fitness interests of collectives, because it allows that particles can partly construct their own social niche.

7.3 Externalism, interactionism and separation of timescales

There is a sense that the distinction drawn in Chapter 3 between interactionist and externalist explanations is timescale-dependent. It seems natural to suppose social niche modifiers - often structural features of the biological world - evolve more slowly than social behaviours (these can be very simple traits). On a long timescale, the evolution of social niche modifiers and of individuality can be understood in an interactionist fashion. However, if we examine the situation on a shorter timescale, it is possible to move the social niche into the 'external' domain (assuming it to be constant over the timescale considered). This would be simply to explain the evolution of the social trait as a straightforward adaptive response to the social niche, in the manner typical of explanations in social evolution theory. What gives social niche construction its novelty is the mixing of timescales.

[3]E.g. Michod (1999, p.63): "Adaptive design for the organism depends upon the regulation of ... lower-level units, and this regulation allows the organism to emerge as a unit of selection, or an individual, it its own right."

7.4 On the sacrifice of individual fitness interests for the greater good

It is widely accepted that ETIs are examples of the evolution of cooperation. A lack of clarity about what this means has led some authors to understand ETIs in terms of loss of fitness for lower level entities and gain of fitness for higher level entities. I discussed one prominent example of this sort of thinking in Chapter 5 on Fitness Export Theory. I now want to examine more generally the idea that cooperation involves individual sacrifice for a greater good. Commenting on Gardner and Grafen (2009)'s use of inclusive fitness theory to explain away apparent altruism as a case of genetic self-interest, Sober and Wilson (2011) complain that "The net result is that any helping behaviour that evolves because of natural selection gets viewed as a form of genetic self-interest. This may seem like a pleasing consequence until it is realized that 'self-interest' has now become an all-encompassing category. When altruism evolves, this is consistent with the heuristic idea of self-interest, as altruists are getting their altruistic genes into the next generation by helping other altruists. The idea that altruism is good for the group but bad for the individual has been lost."

As I have said several times now, the idea that altruism is good for the group but bad for the individual is central to Fitness Export theory: "the essence of an evolutionary transition in individuality is that the lower-level individuals must 'relinquish' their 'claim' to individual fitness in favor of the survival and reproduction of the new higher-level unit" (Michod and Nedelcu, 2003).

I disagree with this widespread intuition that altruism is bad for the individual/particle in cases where it evolves. In cases where altruism evolves, there must be a social niche that favours it. (High relatedness will often do the trick and is what Gardner and Grafer have in mind.) I grant that altruism *would have been* bad for the individual in the counterfactual situation where the prevailing social niche did not favour altruism. But I reject the claim that there is a real, factual loss of fitness - that altruism is 'bad for the individual' in cases where it evolves. To suppose that there is, one needs to selectively ignore some relevant facts while foregrounding others. Specifically, one needs to ignore alignment of fitness interests (most obviously due to inclusive fitness or reliable reciprocity) and fitness gains made by successful collective action, while focusing on the fitness loss that would have been incurred by altruists in a different social niche (one that does not, in fact, obtain)[4].

There is an interesting analogy here with the distinction between individualism and collectivism in political philosophy. For the purposes of drawing the analogy (and to avoid digressions well beyond the scope of this thesis), I define those positions as follows. Political collectivism has it that (at least some) social collectives have moral ends that are ontologically distinct from

[4]There is still value to be had from distinguishing between a) social behaviours of entities that are explicable in terms of an evolutionary process acting at that level and b) social behaviours of entities that are not explicable in terms of an evolutionary process acting at that level - these two cases call for different *causal* stories (Okasha, 2015) but they do not alter the facts about individual self-sacrifice.

the moral ends of the individual people that make them up while individualism has it that only individuals are moral ends (Triandis, 1995)[5]. Proponents of collectivist views see individual self-interest as a selfish and destructive force that must be reined-in by state action, lest it harm the well-being of society as a whole. Proponents of individualist views see individual self-interest as a creative force that, if permitted, gives rise to division of labour, efficient resource allocation and increased productivity - yielding greater well-being at the societal level. It is not my business to offer judgements on political philosophy. However, I wonder if the shadow of the collectivist view exerts its influence over those who think evolutionary transitions involve self-sacrificial behaviour on the part of the entities in the role of particles? Popper (1945, p.101) observes that the intuition that collective well-being entails actual individual sacrifice is very widespread. He traces the running-together of self-interest and destructive selfishness back to Plato. As Popper argues, not all authors agree that pursuit of individual self-interest is a selfish and destructive force. (Recall that collective action problems of the kind discussed in this thesis are not zero sum games.) Adam Smith famously argued that the pursuit of individual self-interest can give rise to social benefit as unintended consequence and can itself be seen as a form of mutual cooperation (I would add: but never as a form of strong altruism). I suggest that major evolutionary transitions can be better understood as unintended consequences of the pursuit of self-interest, in something close to Smith's sense.

7.5 Selection-for social niche construction is selection-of evolutionary transition

SNC does not invoke a collective-level evolutionary process to explain how collectives became units of evolution (Section 7.2). Nor does it assume that particle-level selection must always be opposed to the fitness interests of collectives, because it allows that particles can partly construct their own social niche (Section 7.4). Social niche construction explains the advent of a collective-level evolutionary process as a side-effect of a particle-level evolutionary process that aligns the fitness interests of particles within collectives as a means to making the particles fitter (Section 3.9).

The claim that social niche construction explains the onset of evolutionary transitions as a side-effect of something else warrants elaboration. Sober (1984) draws a distinction that will be useful here between *selection-for* a trait and *selection-of* a trait. The distinction is intended to draw attention to the difference between causes and effects: selection-for describes the cause of

[5]See also von Mises (1949, Chapter 8):"According to the [doctrine of collectivism], society is an entity living its own life, independent of and separate from the lives of the various individuals, acting on its own behalf and aiming at its own ends which are different from the ends sought by the individuals. Then, of course, **an antagonism between the aims of society and those of its members can emerge**. In order to safeguard the flowering and further development of society it becomes necessary to master the selfishness of the individuals and to compel them to sacrifice their egoistic designs to the benefit of society." (Emphasis is mine.) Clearly there is an analogy here with the distinction between particle fitness (and collective fitness_1, an of aggregate particle fitness) on the one hand - and collective fitness_2 on the other (where fitness_2 is ontological distinct from particle fitness - neither proportional to it nor defined in terms of it). That distinction is central to Fitness Export theory.

a selection process, while selection-of refers to its effects. Suppose there has been selection-for blood capable of transporting oxygen. Say it does this using haemoglobin (which happens to impart red colouration). In this example there has been selection-for blood capable of oxygen-transport and selection-of blood that is red. Sober's example involves the evolution of the human jaw. Suppose palaeontologists tell us there was selection-for a jaw with a certain range of functions. That just means that jaws with those functions gave higher fitness to their bearers than those without them. At the same time, there was selection-of the chin - not because chins are adaptive in themselves but because they are side-effects of something else that is adaptive. Sober's distinction matters here because I propose that, in the early stages of an ETI, there is *selection-of* upshifting the level of selection - but not *selection-for* it[6].

Selection-of can occur due to genetic factors (such as linkage or pleiotropy) or due to phenotypic factors (such as developmental constraint or simple side-effects, i.e. spandrels)[7]. It is the phenotypic version of selection-of that I am invoking here when I say there is selection-of upshifting the level of selection - but not selection-for it. To be clear: I do not wish to claim that there is a genetic basis for collective-level selection that is distinct from the genetic basis for social niche construction and that the former hitchhikes on the selective success of the latter due to close linkage. Nor do I wish to claim that the genetic basis of social niche construction actually codes for two distinct phenotypic traits - one being social niche construction and the other being collective-level selection (that would be pleiotropy). What I do wish to claim is that the phenomenon of social niche construction *is* the phenomenon of evolutionary transition, given the right circumstances. No higher-level selection is invoked as a cause, even though it comes about later, as an effect (and can then be used subsequently, to explain further adaptations at the higher level, after suppression of lower-level selection has been attained)

Once a collective-level evolutionary process has become established, the alignment of fitness interests within collectives (I'm focusing here on fraternal evolutionary transitions, i.e. the coming together of like-kinds) enables the reproductive and functional division of labour. Once that door has been opened, the scope for adaptation of complex form and function at collective level is endless, as witnessed by the phenotypic diversity of the fungi, animals, land plants, brown algae and red algae.

[6]Granted, once collective-level selection is established then collective-level adaptations may arise that both raise collective fitness and further cement the status of collectives as units of evolution. But the thing to be explained here is how the collective-level evolutionary process gets started in the first place.

[7]Sober's discussion of pleiotropy (1984, p.101) appears to run together the genetic concept of pleiotropy, where one locus affects two or more separate traits (only one of which is selected-for) with the phenotypic concept of a spandrel, where a trait arises not due to selection-for it but as a side-effect of another trait that is selected-for).

Appendix A

Glossary

Term	Meaning
base game	A game (used in this thesis to model some social niches) that is unmodified. Compare 'effective game'. See Chapters 3 and 4.
bottleneck	A single-celled stage in the life-history of multi-cellular orgasisms (e.g zygote, seed or spore) or its analog. I set these up a 'social niche modifiers' in this thesis. See 4.
Bottleneck Model	A simulation model described in Chapter 4.
collective	Higher level entities in a two-level part-whole hierarchy (see Figure 1.1).
collective reproducers	Entities that reproduce and that are made of parts that also reproduce, along the lines of the collectives and their constituent particles (illustrated in Figure 1.1) that feature prominently in this thesis. A term I borrow from Godfrey-Smith (2009).

collectivism, fallacy of	A position in political philosophy that holds that human collectives have moral ends that are ontologically distinct from (and may conflict with and sometimes take precedence over) the moral ends of the individuals they contain. See Fitness Export Theory.
constructivist	See interactionist explanatory stance.
Darwin's 'one special difficulty'	Darwin worried about how to explain the evolution of cooperation. See Chapter 3.
darwinian individuality	A property predicated of evolutionary units.
Darwinian machine	A system in which there is a population that exhibits heritable variation in fitness, i.e. a system capable of evolutionary change.
diachronic	Something happening over time - in this context, it refers to the way the levels of selection (and nested structure of the part-whole biological hierarchy) are not static but change over deep evolutionary time (Okasha, 2005b). Compare 'synchronic'.
dialectical	See interactionist explanatory stance.
effective game	A game (used in this thesis to model some social niches) that is actually being played. It is the resultant of the base game and any social niche modifiers (e.g. assortment) that obtain. Compare 'base game'. See Chapters 3 and 4.
egalitarian transition	An evolutionary transition in individuality involving unlike kinds of thing coming together to form a new unit (e.g. a lichen arising out of an association between a photobiont (cyanobacterium or alga) and a fungus). Term due to Queller (1997).

ETI	See evolutionary transition in individuality .
evolutionary transition in individuality	Defined, in terms of character-fitness covariances at different hierarchical levels, in Section 1.4.
evolutionary-unit-theory of biological individuality	A loose grouping of theories all of which take an evolutionary approach to biological individuality. See Section 2.3.
externalist explanatory stance	The position that an explanation for the properties of biological entities is best stated mainly in terms of their adaptation to factors external to those entities; common in evolutionary biology (in the guise of adaptationism). See Section 3.2.
Fitness Decoupling Theory	Another name for Fitness Export Theory.
Fitness Export Theory	The currently best-accepted theory about how evolutionary transitions in individuality evolve, Chapter 5.
fitness_1	A fitness measure used in some abstract models of multilevel selection. A measure of how many particles a focal collective contributes to the global particle population. See Section 2.4.
fitness_2	A fitness measure used in some abstract models of multilevel selection. A measure of how many offspring collectives a focal collective contributes to the global collective population. See Section 2.4. I discuss its relevance to ETIs in Chapter 6.

fraternal transition	An evolutionary transition in individuality involving like kinds of thing coming together to form a new unit (e.g. multicellular plants evolving from single-celled ones). Term due to Queller (1997). Compare egalitarian transition.
group	See collective.
interactionist explanatory stance	The position that an explanation for the properties of biological entities is best stated in terms of the interplay of internal and external factors. See Section 3.2.
internalist explanatory stance	The position that an explanation for the properties of biological entities is best stated mainly in terms of factors internal to those entities; common in developmental biology (in the guise of genetic pre-formationism). See Section 3.2.
level of selection	Hierarchical level at which entity fitness covaries with entity character (Section 1.3).
Lewontin Conditions	A recipe for evolutionary change featuring the ingredients of heritable variation in fitness, Section 1.3.
life-history bottleneck	See bottleneck.
major evolutionary transitions	Events in the history of life on earth involving a coalescing or coming-together of entities over evolutionary time, yielding new levels of biological individuality. See evolutionary transition in individuality.
Major Evolutionary Transitions research programme	A scientific research programme concerned with explaining major evolutionary transitions
Major Transitions	See Major Evolutionary Transitions research programme

MLS1,MLS2	MLS1 and MLS2 are each a class of models of multi-level selection, see Section 2.4.
one special difficulty	See Darwin.
particle	Lower level entities in a two-level part-whole hierarchy (see Figure 1.1).
Price Approach	Relative strength of character-fitness covariance at each of two levels can be compared quantitatively in some circumstances, see Section 2.3.3.
Price Equation	A covariance equation to describe evolutionary change, see 2.3.3.
problem of biological individuality	The ancient problem of giving a principled way to define what things are biolgical individuals (and how to tell them apart from their parts and groups of them).
propagule	A small part of a biological individual, emitted by it for the purposes of reproduction. See Chapter 4. .
selection-for/selection-of	There is selection-for a trait when the trait causes its bearers to be fitter than its non-bearers; there is selection-of a trait when the trait increases in frequency for some other reason, Sober's (1984, p.97) terminology. See Section 3.9.
selection-of	See selection-for/selection-of.
SNC	See social niche construction.
SNM	See social niche modifier.

social behaviour	Any behaviour that affects the fitness of members of the population other than the actor, Section 3.5.
social niche	The selective context in which social behaviour occurs, affecting the strength and direction of selection on it. In game theoretic terms, the social niche is the effective game being played, once all relevant factors have been taken into account. (Relevant factors include the underlying game, any social niche modifiers (SNMs) that may be present and the frequencies of various social traits in the population).
social niche construction	A circular process in which organisms modify their own social niche in such a way as to influence the conditions of their own social evolution.
social niche modifier (SNM)	A structural feature of the biological world that has the function of modifying a social niche. It is a trait that alters the effective game being played by its bearers, causing it to differ from the counterfactual game they would have been playing if the social niche modifier had not acted. Examples include factors such as population structure, bottlenecks, relatedness, punishment and worker-policing. See Table 3.1 on page 38 for many examples that make the concept clear. And see Section 3.5.
Social Niche Modifying trait	In one of my models, a trait that acts as a social niche modifier (see social niche modifier).
Social Trait (ST)	In one of my models, a trait that affects the fitness of individuals other than the actor, sometimes having values appropriately labelled ‘cooperate’ or ‘defect’ .
ST	See social trait.

synchronic	A snapshot of a situation at a point in time (sensu Okasha, 2005b). E.g. the debate between those who think genes are the real level of selection and those who thinks organisms are the real level of selection is synchronic in the sense that it pre-supposes the existence of those levels of organisation and asks about which one is *now* the real level. Synchronic approaches to biological individuality neglect or sideline the fact that those levels haven't always been there but themselves need explanation - compare 'diachronic'.
T-S plane	A compression of the space of two-player two-stategy symmetric games, due to Santos et al. (2006b), see Section 3.8.1. Useful for thinking about collective action between like-kinds (and fraternal transitions).
unit of evolution	The type of unit for which entity fitness covaries with mean character of entity's offspring. Units of evolution show selection and heredity. (See Section 1.3).
unit of selection	The type of unit found at the level of selection. See Section 1.3, (abstractly: possibly particles or collectives; concretely: possibly genes, chromosomes, cells, multicellulars organisms or eusocial colonies). The term is used imprecisely in the literature. I have tried to be more precise in this thesis - see Section 1.3.
VV-fitness	A fitness concept inspired by Van Valen. I claim that it makes a lot of sense and can overcome several problems current theories about evolutionary transitions cannot. See Chapter 6.
Weismannian organism	An idealization 1. that is central to the modern evolutionary synthesis and 2. that is applicable to a tiny proportion of actual organisms, and 3. where early germline segregation means acquired characters are not heritable and intra-organismic evolution is negligible. See Section 2.3.

Bibliography

Abrams, M. (2009). The unity of fitness. *Philosophy of Science*, 76(5):750–761.

Abrams, M. (2012). Measured, modeled, and causal conceptions of fitness. *Frontiers in Genetics*, 3.

Akçay, E. and Roughgarden, J. (2011). The evolution of payoff matrices: providing incentives to cooperate. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1715):2198–2206.

Anderson, C. and McShea, D. W. (2001). Individual versus social complexity, with particular reference to ant colonies. *Biological Reviews of the Cambridge Philosophical Society*, 76(02):211–237.

Anderson, O. R. (2013). *Comparative Protozoology: Ecology, Physiology, Life History*. Springer.

Archetti, M. (2009). Cooperation as a volunteer's dilemma and the strategy of conflict in public goods games. *Journal of Evolutionary Biology*, 22(11):2192–2200.

Archetti, M. and Scheuring, I. (2012). Review: game theory of public goods in one-shot social dilemmas without assortment. *Journal of Theoretical Biology*, 299:9–20.

Archetti, M., Úbeda, F., Fudenberg, D., Green, J., Pierce, N. E., and Douglas, W. Y. (2011). Let the right one in: a microeconomic approach to partner choice in mutualisms. *The American Naturalist*, 177(1):75–85.

Ariew, A. and Lewontin, R. C. (2004). The confusions of fitness. *British Journal for the Philosophy of Science*, pages 347–363.

Aviles, L. (2000). Nomadic behaviour and colony fission in a cooperative spider: life history evolution at the level of the colony? *Biological Journal of the Linnean Society*, 70(2):325–339.

Avise, J. C. (1991). Ten unorthodox perspectives on evolution prompted by comparative population genetic findings on mitochondrial DNA. *Annual Review of Genetics*, 25(1):45–69.

Axelrod, R. and Hamilton, W. (1981). The evolution of cooperation. *Science*, 211(4489):1390.

Ayre, D. J. and Grosberg, R. K. (2005). Behind anemone lines: factors affecting division of labour in the social cnidarian anthopleura elegantissima. *Animal behaviour*, 70(1):97–110.

Barberousse, A., Morange, M., and Pradeu, T. (2009). *Mapping the future of biology: evolving concepts and theories*. Springer.

Beatty, J. (1992). Fitness: Theoretical contexts. In Keller, E. F. and Lloyd, E. A., editors, *Keywords in Evolutionary Biology*, pages 115–119. Harvard University Press.

Beklemishev, W. (1969). *Principles of Comparative Anatomy of Invertebrates: Promorphology*. Oliver and Boyd.

Bennett, N. C. and Faulkes, C. G. (2000). *African mole-rats: ecology and eusociality*. Cambridge University Press.

Bergstrom, C. T., Bronstein, R. J. L., Bshary, R., Connor, R. C., Daly, M., Frank, S. A., Gintis, H., Keller, L., Leimar, O., Noë, R., et al. (2003). Interspecific mutualism: Puzzles and predictions. In *Genetic and cultural evolution of cooperation*. MIT Press.

Bergstrom, C. T. and Pritchard, J. (1998). Germline bottlenecks and the evolutionary maintenance of mitochondrial genomes. *Genetics*, 149(4):2135–2146.

Birky, C. W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. *Proceedings of the National Academy of Sciences*, 92(25):11331–11338.

Bonner, J. T. (1965). *Size and Cycle - An essay on the structure of biology*. Princeton University Press.

Bonner, J. T. (1974). *On Development: The biology of form*. Harvard University Press.

Bonner, J. T. (2000). *First Signals - The Evolution of Multicellular Development*. Princeton University Press.

Boomsma, J. J. (2009). Lifetime monogamy and the evolution of eusociality. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 364(1533):3191–207.

Borenstein, E., Kendal, J., and Feldman, M. (2006). Cultural niche construction in a metapopulation. *Theoretical population biology*, 70(1):92–104.

Bossert, W., Qi, C. X., and Weymark, J. A. (2013). Extensive social choice and the measurement of group fitness in biological hierarchies. *Biology & Philosophy*, 28(1):75–98.

Bouchard, F. (2008). Causal processes, fitness, and the differential persistence of lineages. *Philosophy of Science*, 75(5):560–570.

Bouchard, F. (2013). What is a symbiotic superindividual and how do you measure its fitness. In Bouchard, F. and Huneman, P., editors, *From groups to individuals: evolution and emerging individuality*. MIT Press.

Bourke, A. (2011). *Principles of Social Evolution*. Oxford University Press.

Bourke, A. and Franks, N. (1995). *Social Evolution In Ants*. Princeton University Press.

Bourrat, P. (2015a). Levels of selection are artefacts of different fitness temporal measures. *Ratio*, 28(1):40–50.

Bourrat, P. (2015b). Levels, time and fitness in evolutionary transitions in individuality. *Philosophy & Theory in Biology*, 7.

Boyd, R., Gintis, H., and Bowles, S. (2010). Coordinated punishment of defectors sustains cooperation and can proliferate when rare. *Science*, 328(5978):617–620.

Brandon, R. N. (1978). Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A*, 9(3):181–206.

Brandon, R. N. (1990). *Adaptation and Environment*. Princeton University Press.

Braude, S. (2000). Dispersal and new colony formation in wild naked mole-rats: evidenceagainst inbreeding as the system of mating. *Behavioral Ecology*, 11(1):7–12.

Breden, F. and Wade, M. J. (1991). Runaway social evolution: Reinforcing selection for inbreeding and altruism. *Journal of Theoretical Biology*, 153(3):323–337.

Brett, R. (1991). The population structure of naked mole-rat colonies. In Sherman, P. W., Jarvis, J. U., and Alexander, R. D., editors, *The Biology of the Naked Mole-Rat*, pages 97–136. Princeton University Press.

Brown, G. W. (1947). On small-sample estimation. *The Annals of Mathematical Statistics*, pages 582–585.

Burnet, F. (1969). *Self and Not-Self: Cellular Immunology*. Cambridge University Press.

Burnett, J. H. (2003). *Fungal Populations and Species*. Oxford University Press.

Burt, A. and Trivers, R. (2006). *Genes in conflict: the biology of selfish genetic elements*. Harvard University Press.

Buss, L. (1987). *The Evolution of Individuality*. Princeton University Press.

Buss, L. W. (1983). Evolution, development, and the units of selection. *Proceedings of the National Academy of Sciences*, 80(March):1387–1391.

Byerly, H. C. and Michod, R. E. (1991). Fitness and evolutionary explanation. *Biology and Philosophy*, 6(1):1–22.

Calcott, B. (2008). Assessing the fitness landscape revolution. *Biology and Philosophy*, 23(5):639–657.

Calcott, B. (2011). Alternative patterns of explanation for major transitions. In Calcott, B. and Sterelny, K., editors, *The major transitions in evolution revisited*. MIT Press.

Cao, L., Ohtsuki, H., Wang, B., and Aihara, K. (2011). Evolution of cooperation on adaptively weighted networks. *Journal of theoretical biology*, 272(1):8–15.

Cartwright, N. (1999). *The Dappled World: A study of the boundaries of science*. Cambridge University Press.

Chao, L. and Levin, B. R. (1981). Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences*, 78(10):6324–6328.

Charlesworth, B. (1980). *Evolution in Age-Structured Populations*. Cambridge University Press.

Charlesworth, D., Morgan, M., and Charlesworth, B. (1993). Mutation accumulation in finite outbreeding and inbreeding populations. *Genetical Research*, 61(01):39–56.

Charrier, B., Coelho, S. M., Le Bail, A., Tonon, T., Michel, G., Potin, P., Kloareg, B., Boyen, C., Peters, A. F., and Cock, J. M. (2008). Development and physiology of the brown alga ectocarpus siliculosus: two centuries of research. *New Phytologist*, 177(2):319–332.

Ciszek, D. (2000). New colony formation in the 'highly inbred' eusocial naked mole-rat: outbreeding is preferred. *Behavioral Ecology*, 11(1):1–6.

Clarke, E. (2010). *Biological Individuality and the Levels of Selection*. PhD thesis, University of Bristol.

Clarke, E. (2011a). Plant Individuality and Multilevel Selection Theory. In *The Major Transitions in Evolution Revisited*, chapter 11, pages 227–250. MIT Press, Cambridge, MA.

Clarke, E. (2011b). The Problem of Biological Individuality. *Biological Theory*, 5(4):1–15.

Clarke, E. (2012). Plant individuality: a solution to the demographer's dilemma. *Biology & Philosophy*, 27(3):321–361.

Clarke, E. (2013). The multiple realizability of biological individuals. *Journal of Philosophy*, 8.

Clarke, E. (2014). Origins of evolutionary transitions. *Journal of biosciences*, 39(2):303–317.

Clarke, E. and Okasha, S. (2013). Species and organisms: What are the problems? In *From Groups to Individuals: Evolution and Emerging Individuality*. MIT Press.

Clay, K. (1990). Fungal endophytes of grasses. *Annual Review of Ecology and Systematics*, 21:275–297.

Clayton, M. (1988). Evolution and life histories of brown algae. *Botanica Marina*, 31(5):379–388.

Coelho, S. M., Peters, A. F., Charrier, B., Roze, D., Destombe, C., Valero, M., and Cock, J. M. (2007). Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene*, 406(1):152–170.

Cohen, G. (1978). *Karl Marx's Theory of History: A Defence*. Oxford University Press.

Colman, A. M. (2013). *Game theory and its applications: In the social and biological sciences*. Psychology Press.

Connor, R. C. (1992). Egg-trading in simultaneous hermaphrodites: an alternative to tit-for-tat. *Journal of Evolutionary Biology*, 5(3):523–528.

Crespi, B. J. (2001). The evolution of social behavior in microorganisms. *Trends in Ecology & Evolution*, 16(4):178–183.

Crespi, B. J. (2004). Vicious circles: positive feedback in major evolutionary and ecological transitions. *Trends in Ecology & Evolution*, 19(12):627–633.

Damuth, J. and Heisler, I. L. (1988). Alternative formulations of multilevel selection. *Biology and Philosophy*, 3:407–430.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology*, 31(1):169–193.

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

Dawkins, R. (1982). *The extended phenotype: the gene as the unit of selection*. Oxford University Press, Oxford.

De Monte, S. and Rainey, P. B. (2014). Nascent multicellular life and the emergence of individuality. *Journal of Biosciences*, 39(2):237–248.

De Sousa, R. (2005). Biological individuality. *Croatian Journal of Philosophy*, 5(14):195–218.

Degennaro, F. and Weller, S. (1984). Growth and reproductive characteristics of field bindweed (convolvulus arvensis) biotypes. *Weed Science*, pages 525–528.

Diaz Ricci, J. C. and Hernández, M. E. (2000). Plasmid effects on escherichia coli metabolism. *Critical reviews in biotechnology*, 20(2):79–108.

Dobzhansky, T. (1970). *Genetics of the Evolutionary Process*. Columbia University Press.

Dobzhansky, T. G. (1937). *Genetics and the Origin of Species*. Columbia University Press.

Douglas, A. (1989). Mycetocyte symbiosis in insects. *Biological Reviews*, 64(4):409–434.

Dunn, C. (2009). Siphonophores. *Current Biology*, 19(6):R233–R234.

Dunn, C. W. (2005). Complex colony-level organization of the deep-sea siphonophore bargmannia elongata (cnidaria, hydrozoa) is directionally asymmetric and arises by the subdivision of pro-buds. *Developmental Dynamics*, 234(4):835–845.

Dunn, C. W. and Wagner, G. P. (2006). The evolution of colony-level development in the Siphonophora (Cnidaria:Hydrozoa). *Development Genes and Evolution*, 216(12):743–54.

Dupré, J. (1995). *The disorder of things: Metaphysical foundations of the disunity of science*. Harvard University Press.

Dupré, J. (2010). The polygenomic organism. *Sociological Review*, 58:19–31.

Emson, R. and Wilkie, I. (1980). *Fission and autotomy in echinoderms*. Aberdeen University Press.

Eshel, I. and Cavalli-Sforza, L. L. (1982). Assortment of encounters and evolution of cooperativeness. *Proceedings of the National Academy of Sciences*, 79(4):1331–1335.

Ewald, P. W. (1987). Transmission modes and evolution of the parasitism-mutualism continuum. *Annals of the New York Academy of Sciences*, 503(1):295–306.

Farrant, P. (1985). Reproduction in the temperate australian soft coral capnella gaboensis. In *Proceedings of the 5th International Coral Reef Congress*, volume 4, pages 319–324.

Faulkes, C., Abbott, D., and Mellor, A. (1990). Investigation of genetic diversity in wild colonies of naked mole-rats (Heterocephalus glaber) by DNA fingerprinting. *Journal of Zoology*, 221(1):87–97.

Faulkes, C., Bennett, N., Bruford, M. W., O'brien, H., Aguilar, G., and Jarvis, J. (1997). Ecological constraints drive social evolution in the african mole–rats. *Proceedings of the Royal Society of London B: Biological Sciences*, 264(1388):1619–1627.

Fisher, R. (1930). *The Genetical Theory of Natural Selection.* Clarendon Press.

Fisher, R. M., Cornwallis, C. K., and West, S. A. (2013). Group formation, relatedness, and the evolution of multicellularity. *Current Biology*, 23:1120–1125.

Fletcher, J. A. and Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. *Proceedings of the Royal Society B: Biological Sciences*, 276(1654):13–19.

Fletcher, J. A. and Zwick, M. (2004). Strong altruism can evolve in randomly formed groups. *Journal of Theoretical Biology*, 228(3):303–13.

Folse, H. (2011). *Evolution and Individuality: Beyond the Genetically Homogenous Organism*. PhD thesis, Stanford University.

Folse, H. J. and Roughgarden, J. (2010). What is an individual organism? a multilevel selection perspective. *The Quarterly Review of Biology*, 85(4):447–472.

Fox, G. A. (1990). Perennation and the persistence of annual life histories. *American Naturalist*, pages 829–840.

Francis, L. (1979). Contrast between solitary and clonal lifestyles in the sea anemone anthopleura elegantissima. *American Zoologist*, 19(3):669–681.

Frank, S. A. (2012). Natural selection. IV. The Price equation. *Journal of Evolutionary Biology*, 25(6):1002–1019.

Fu, F., Nowak, M., and Hauert, C. (2010). Invasion and expansion of cooperators in lattice populations: Prisoner's dilemma vs. Snowdrift games. *Journal of theoretical biology*, 266:358–366.

Gardner, A. (2008). The Price Equation. *Current Biology*, 18(5):R198–R202.

Gardner, A. (2009). Adaptation as organism design. *Biology Letters*, 5(6):861–864.

Gardner, A. (2013). Adaptation of individuals and groups. In Bouchard and Huneman, editors, *From groups to individuals: evolution and emerging individuality*. MIT Press.

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 West, S. A. (2010). Greenbeards. *Evolution*, 64(1):25–38.

Geller, J. B. and Walton, E. D. (2001). Breaking up and getting together: evolution of symbiosis and cloning by fission in sea anemones (genus anthopleura). *Evolution*, 55(9):1781–1794.

Gillespie, J. H. (1974). Natural selection for within-generation variance in offspring number. *Genetics*, 76(3):601–606.

Gillespie, J. H. (1977). Natural selection for variances in offspring numbers: a new evolutionary principle. *The American Naturalist*, 111(981):1010–1014.

Gillespie, J. H. (1998). *Population Genetics: a concise guide*. Johns Hopkins University Press.

Glass, N. L., Jacobson, D. J., and Shiu, P. K. (2000). The genetics of hyphal fusion and vegetative incompatibility in filamentous ascomycete fungi. *Annual Review of Genetics*, 34(1):165–186.

Godfrey-Smith, P. (1996). *Complexity and the Function of Mind in Nature*. Cambridge University Press.

Godfrey-Smith, P. (2001). Organism, environment, and dialectics. In Singh, R., Krimbas, C., Paul, D., and Beatty, J., editors, *Thinking about evolution: Historical, philosophical, and political perspectives*, pages 253–266. Cambridge University Press.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology & Philosophy*, 21(5):725–740.

Godfrey-Smith, P. (2008). Varieties of Population Structure and the Levels of Selection. *The British Journal for the Philosophy of Science*, 59(1):25–50.

Godfrey-Smith, P. (2009). *Darwinian Populations and Natural Selection*. Oxford University Press.

Godfrey-Smith, P. (2011a). Agents and acacias: replies to Dennett, Sterelny, and Queller. *Biology & Philosophy*, 26(4):501–515.

Godfrey-Smith, P. (2011b). Darwinian populations and transitions in individuality. In *The Major Transitions in Evolution Revisited*. MIT Press.

Godfrey-Smith, P. (2014). *Philosophy of Biology*. Princeton University Press.

Goodnight, C. J., Schwartz, J. M., and Stevens, L. (1992). Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *American Naturalist*, pages 743–761.

Gore, J., Youk, H., and Van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature*, 459(7244):253–256.

Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Harvard University Press.

Gould, S. J. and Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, 205(1161):581–598.

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

Greig, D. and Travisano, M. (2004). The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proceedings of the Royal Society of London B: Biological Sciences*, 271(Suppl 3):S25–S26.

Griesemer, J. (2000). The units of evolutionary transition. *Selection*, 1:67–80.

Griesemer, J. and Wimsatt, W. (1989). Picturing Weismannism: A Case Study of Conceptual Evolution. In Ruse, M., editor, *What the Philosophy of Biology Is*, pages 75–138. Springer.

Grosberg, R. K. (1988). The Evolution of Allorecognition Specificity in Clonal Invertebrates. *Quarterly Review of Biologyr*, 63(4):377–412.

Grosberg, R. K. and Strathmann, R. R. (1998). Once cell, two cell, red cell, blue cell: the persistence of a unicellular stage in multicellular life histories. *Trends in ecology & evolution*, 13(3).

Grosberg, R. K. and Strathmann, R. R. (2007). The Evolution of Multicellularity: A Minor Major Transition? *Annual Review of Ecology, Evolution, and Systematics*, 38(1):621–654.

Haig, D. and Grafen, A. (1991). Genetic scrambling as a defence against meiotic drive. *Journal of Theoretical Biology*, 153:531–558.

Hamilton, W. (1964). The genetical evolution of social behaviour. II. *Journal of Theoretical Biology*, 7(1):17–52.

Hamilton, W. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In *Biosocial Anthropology*. Malaby Press, London.

Hammond, P. J. (2005). Utility invariance in non-cooperative games. In U., S. and Traub, S., editors, *Advances in Public Economics: Utility, choice and welfare*, pages 31–50. Springer.

Hanschen, E. R., Shelton, D. E., and Michod, R. E. (2015). Evolutionary transitions in individuality and recent models of multicellularity. In Ruiz-Trillo, I. and Nedelcu, A., editors, *Evolutionary Transitions to Multicellular Life*, pages 165–188. Springer.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162(3859):1243–8.

Harper, J. L. (1977). *Population Biology of Plants.* Academic Press.

Harrison, E., Koufopanou, V., Burt, A., and MacLean, R. (2012). The cost of copy number in a selfish genetic element: the 2-μm plasmid of saccharomyces cerevisiae. *Journal of evolutionary biology*, 25(11):2348–2356.

Hartl, D. L. and Clark, A. G. (1997). *Principles of Population Genetics*. Sinauer Associates.

Hedrick, P. (2011). *Genetics of Populations*. Jones & Bartlett Learning.

Heisler, I. L. and Damuth, J. (1987). A method for analyzing selection in hierarchically structured populations. *American Naturalist*, pages 582–602.

Herron, M. D., Hackett, J. D., Aylward, F. O., and Michod, R. E. (2009). Triassic origin and early radiation of multicellular volvocine algae. *Proceedings of the National Academy of Sciences*, 106(9):3254–3258.

Herron, M. D. and Michod, R. E. (2008). Evolution of complexity in the volvocine algae: transitions in individuality through darwin's eye. *Evolution*, 62(2):436–451.

Herron, M. D., Rashidi, A., Shelton, D. E., and Driscoll, W. W. (2013). Cellular differentiation and individuality in the minor multicellular taxa. *Biological Reviews*, 88(4):844–861.

Honeycutt, R., Nelson, K., Schlitter, D., and Sherman, P. (1991). Genetic variation within and among populations of the naked mole-rat: evidence from nuclear and mitochondrial genomes. In Sherman, P. W., Jarvis, J. U., and Alexander, R. D., editors, *The Biology of the Naked Mole-Rat*, pages 195–208. Princeton University Press.

Hull, D. L. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, 11:311–332.

Huneman, P. (2013). Adaptations in transitions: How to make sense of adaptation when beneficiaries emerge simultaneously with benefits? In Bouchard, F. and Huneman, P., editors, *From Groups to Individuals: Evolution and Emerging Individuality*. MIT press.

Huxley, J. (1912). *The Individual in the Animal Kingdom*. Cambridge University Press.

Ingram, C. M., Troendle, N. J., Gill, C. A., Braude, S., and Honeycutt, R. L. (2015). Challenging the inbreeding hypothesis in a eusocial mammal: population genetics of the naked mole-rat, heterocephalus glaber. *Molecular Ecology*, 24(19):4848–4865.

J.A. Wilson (1999). *Biological Individuality: The Identity and Persistence of Living Entities*. Cambridge University Press.

J.A. Wilson (2000). Ontological Butchery : Organism Concepts and Biological Generalizations. *Philosophy of Science*, 67:S301–S311.

Jackson, A. (2015). *Metagames: The Evolution of Game-Changing Traits*. PhD thesis, University of Southampton.

Jackson, A. and Watson, R. A. (2016). Metagames: A Formal Framework for the Evolution of Game-Changing Behaviours. *Manuscript submitted for publication at the Journal of Theoretical Biology*.

Jackson, J. B. C. and Coates, a. G. (1986). Life Cycles and Evolution of Clonal (Modular) Animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 313(1159):7–22.

Jacob, F. (1970). *The Logic of Life - a History of Heredity*. Pantheon Books.

Janzen, D. (1977). What are dandelions and aphids? *American Naturalist*, 111:586–589.

Jarvis, J. (1981). Eusociality in a mammal: cooperative breeding in naked mole-rat colonies. *Science*, 212(4494):571–573.

Jarvis, J. (1991). Reproduction of naked mole-rats. In *The Biology of the Naked Mole-Rat*, pages 384–425. Princeton University Press.

Jarvis, J. and Bennett, N. (1993). Eusociality has evolved independently in two genera of bathyergid mole-rats but occurs in no other subterranean mammal. *Behavioral Ecology and Sociobiology*, 33(4):253–260.

Ji, Q. and Xian-Jia, W. (2011). Evolutionary games in a generalized Moran Process with arbitrary selection strength and mutation. *Chinese Physics B*, 20(3):030203.

Johnston, I. G., Burgstaller, J. P., Havlicek, V., Kolbe, T., Rülicke, T., Brem, G., Poulton, J., and Jones, N. S. (2015). Stochastic modelling, Bayesian inference, and new in vivo measurements elucidate the debated mtDNA bottleneck mechanism. *eLife*, 4:e07464.

Keller, L., editor (1999). *Levels of selection in evolution*. Princeton University Press.

Kentzoglanakis, K., López, D. G., Brown, S. P., and Goldstein, R. A. (2013). The evolution of collective restraint: policing and obedience among non-conjugative plasmids. *PLoS Comput Biol*, 9:e1003036.

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., Fellbaum, C. R., Kowalchuk, G. A., Hart, M. M., Bago, A., et al. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*, 333(6044):880–882.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.

Kimura, M. et al. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626.

King, N. (2004). The unicellular ancestry of animal development. *Developmental cell*, 7(3):313–325.

Kirby, D. A., Muse, S. V., and Stephan, W. (1995). Maintenance of pre-mRNA secondary structure by epistatic selection. *Proceedings of the National Academy of Sciences*, 92(20):9047–9051.

Kirk, D. L. (2005). *Volvox: a search for the molecular and genetic origins of multicellularity and cellular differentiation*. Cambridge University Press.

Kitcher, P. (1985). *Vaulting Ambition: Sociobiology and the quest for human nature*. MIT Press.

Klekowski, E. J. (1988). *Mutation, Developmental Selection, and Plant Evolution*. Columbia University Press.

Knoll, A. H. (2011). The multiple origins of complex multicellularity. *Annual Review of Earth and Planetary Sciences*, 39:217–239.

Krimbas, C. B. (2004). On fitness. *Biology and Philosophy*, 19(2):185–203.

Kropotkin, P. (1902). *Mutual Aid: A Factor of Evolution*. Heinemann.

Krupp, D. (forthcoming). Causality and the levels of selection. *Trends In Ecology and Evolution*.

Lacey, E. and Sherman, P. (1991). Social organization of naked mole-rat colonies: evidence for divisions of labor. In *The Biology of the Naked Mole-Rat*. Princeton University Press.

Lack, D. (1947). *Darwin's Finches*. Cambridge University Press.

Laland, K. N., Odling-Smee, J., and Feldman, M. W. (2001). Cultural niche construction and human evolution. *Journal of Evolutionary Biology*, 14(1):22–33.

Laland, K. N. and Sterelny, K. (2006). Perspective: seven reasons (not) to neglect niche construction. *Evolution*, 60(9):1751–1762.

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., and Odling-Smee, J. (2015). The extended evolutionary synthesis: its structure, assumptions and predictions. *Proceedings of the Royal Society B*, 282(1813):20151019.

Leigh, E. (1971). *Adaptation and Diversity: Natural History and the Mathematics ofEvolution*. Freeman, Cooper & Co., San Francisco.

Leigh, E. G. (1991). Genes, bees and ecosystems: the evolution of a common interest among individuals. *Trends in Ecology & Evolution*, 6(8):257–262.

Lenton, T. and Watson, A. J. (2011). *Revolutions That Made The Earth*. Oxford University Press.

Levins, R. (1968). *Evolution in changing environments: some theoretical explorations*. Princeton University Press.

Lewontin, R. (1970). The Units of Selection. *Annual Review of Ecology and Systematics*, 1:1–18.

Lewontin, R. (1982). Organism and environment. In Plotkin, H., editor, *Learning, development, and culture: essays in evolutionary epistemology*. Wiley.

Lewontin, R. (1985). The organism as the subject and object of evolution. In Levins, R. and Lewontin, R., editors, *The Dialectical Biologist*. Harvard University Press.

Lewontin, R. (1991). *Biology as Ideology*. Canadian Broadcasting Corporation.

Lewontin, R. (2000). Foreword to the 2nd edition. In Oyama, S., Taylor, P., Fogel, A., Lickliter, R., Sterelny, K., Smith, K. C., and van der Weele, C., editors, *The Ontogeny of Information - Developmental Systems and Evolution*. Duke University Press.

Lieberman, E., Hauert, C., and Nowak, M. A. (2005). Evolutionary Dynamics on Graphs. *Nature*.

Ligrone, R., Duckett, J. G., and Renzaglia, K. S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. *Annals of Botany*, 109(5):851–871.

Lurling, M. and Beekman, W. (2006). Palmelloids formation in chlamydomonas reinhardtii: defence against rotifer predators? *International Journal of Limnology*, 42(02):65–72.

Lynch, M. (1996). Mutation accumulation in transfer rnas: molecular evidence for muller's ratchet in mitochondrial genomes. *Molecular biology and evolution*, 13(1):209–220.

Mackie, Pugh, and Purcell (1987). Siphonophore biology. In *Advances in Marine Biology 24*. London: Academic Press.

Mackie, G. (1986). From aggregates to integrates: physiological aspects of modularity in colonial animals. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 313(1159):175–196.

Macy, M. W. and Flache, A. (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences*, 99(3):7229–7236.

Margulis, L. (1981). *Symbiosis in cell evolution: Life and its environment on the early earth*. Freeman, San Francisco.

Maturana, H. R. and Varela, F. J. (1981). *Autopoiesis and cognition: The realization of the living*. Springer.

Maynard Smith, J. (1958). *The Theory of Evolution*. Cambridge University Press.

Maynard Smith, J. (1964). Group Selection and Kin Selection. *Nature*, 201:1145–1146.

Maynard Smith, J. (1976). Group Selection. *The Quarterly Review of Biology*, 51:277–283.

Maynard Smith, J. (1987). How to model evolution. In Dupré, J., editor, *The Latest on the Best: Essays on Evolution and Optimality*. MIT Press, Cambridge, MA.

Maynard Smith, J. (1988). Evolutionary progress and the levels of selection. In Nitecki, M., editor, *Evolutionary Progress*, pages 219–230. University of Chicago Press.

Maynard Smith, J., Burian, R., Kauffman, S., Alberch, P., Campbell, J., Goodwin, B., Lande, R., Raup, D., and Wolpert, L. (1985). Developmental constraints and evolution: a perspective from the mountain lake conference on development and evolution. *Quarterly Review of Biology*, 60:265–287.

Maynard Smith, J. and Szathmáry, E. (1995). *The Major Transitions in Evolution*. Oxford University Press.

McElreath, R. and Boyd, R. (2008). *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press.

McFadden, C., Grosberg, R., Cameron, B., Karlton, D., and Secord, D. (1997). Genetic relationships within and between clonal and solitary forms of the sea anemone anthopleura elegantissima revisited: evidence for the existence of two species. *Marine Biology*, 128(1):127–139.

McNemar, Q. (1962). *Psychological Statistics*. Wiley.

Mehdiabadi, N. J., Jack, C. N., Farnham, T. T., Platt, T. G., Kalla, S. E., Shaulsky, G., Queller, D. C., and Strassmann, J. E. (2006). Social evolution: kin preference in a social microbe. *Nature*, 442(7105):881–882.

Mendoza, A. and Franco, M. (1998). Sexual reproduction and clonal growth in reinhardtia gracilis (palmae), an understory tropical palm. *American Journal of Botany*, 85(4):521–521.

Michod, R. (1999). *Darwinian Dynamics: Evolutionary Transitions in Fitness and Individuality*. Princeton University Press.

Michod, R. and Herron, M. (2006). Cooperation and conflict during evolutionary transitions in individuality. *Journal of Evolutionary Biology*, 19(5):1406–1409.

Michod, R. and Nedelcu, A. (2003). On the reorganization of fitness during evolutionary transitions in individuality. *Integrative and Comparative Biology*, 43(1):64–73.

Michod, R. and Sanderson, M. (1985). Behavioural structure and the evolution of cooperation. In Greenwood, P., Harvey, P., and Slatkin, M., editors, *Evolution: Essays in Honor of John Maynard Smith*, pages 95–106. Cambridge University Press.

Michod, R. E. (2005). On the transfer of fitness from the cell to the multicellular organism. *Biology and Philosophy*, 20(5):967–987.

Michod, R. E. (2006). The group covariance effect and fitness trade-offs during evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*, 103(24):9113–7.

Michod, R. E. (2011). Evolutionary transitions in individuality: multicellularity and sex. In Calcott, B. and Sterelny, K., editors, *The Major Transitions in Evolution Revisited*, pages 169–197. MIT Press.

Michod, R. E., Nedelcu, A. M., and Roze, D. (2003). Cooperation and conflict in the evolution of individuality. IV. Conflict mediation and evolvability in Volvox carteri. *Bio Systems*, 69(2-3):95–114.

Michod, R. E. and Roze, D. (1999). Cooperation and conflict in the evolution of individuality. III. Transitions in the unit of fitness. *Lectures on Mathematics in the Life Sciences*, pages 47–92.

Michod, R. E., Viossat, Y., Solari, C. A., Hurand, M., and Nedelcu, A. M. (2006). Life-history evolution and the origin of multicellularity. *Journal of Theoretical Biology*, 239(2):257–272.

Mill, J. S. (1848). *Principles of Political Economy with Some of their Applications to Social Philosophy*. London: John W. Parker.

Mills, S. K. and Beatty, J. H. (1979). The propensity interpretation of fitness. *Philosophy of Science*, pages 263–286.

Mills, S. K. and Beatty, J. H. (1994). The propensity interpretation of fitness. In Sober, E., editor, *Conceptual Issues in Evolutionary Biology*, pages 3–23. MIT Press, 2 edition.

Moran, P. A. P. (1958). Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 54, pages 60–71. Cambridge University Press.

Muller, H. J. (1932). Some genetic aspects of sex. *American Naturalist*, pages 118–138.

Nakaya, F., Saito, Y., and Motokawa, T. (2003). Switching of metabolic–rate scaling between allometry and isometry in colonial ascidians. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(1520):1105–1113.

Neher, R. A. and Shraiman, B. I. (2009). Competition between recombination and epistasis can cause a transition from allele to genotype selection. *Proceedings of the National Academy of Sciences*.

Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, pages 1–10.

Nei, M. and Roychoudhury, A. (1974). Sampling variances of heterozygosity and genetic distance. *Genetics*, 76(2):379–390.

Niklas, K. J. (2014). The evolutionary-developmental origins of multicellularity. *American Journal of Botany*, 101(1):6–25.

Niklas, K. J. and Newman, S. A. (2013). The origins of multicellular organisms. *Evolution & development*, 15(1):41–52.

Noë, R. (2001). Biological markets: partner choice as the driving force behind the evolution of mutualisms. In Noë, R., Van Hoof, J., and Hammer, editors, *Economics in Nature*, pages 93–118. Cambridge University Press.

Nowak, M. (2006). *Evolutionary dynamics: exploring the equations of life*. Harvard University Press.

Odling-Smee, F. J., Laland, K. N., and Feldman, M. W. (2003). *Niche construction: the neglected process in evolution*. Princeton University Press.

Ohtsuki, H., Hauert, C., Lieberman, E., and Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs. *Nature*, 441:502–505.

Okasha, S. (2005a). Maynard Smith on the levels of selection question. *Biology and Philosophy*, 20(5):989–1010.

Okasha, S. (2005b). Multilevel Selection and the Major Transitions in Evolution. *Philosophy of Science*, 72(5):1013–1025.

Okasha, S. (2006). *Evolution and the Levels of Selection*. Oxford University Press.

Okasha, S. (2009). Individuals, groups, fitness and utility: multi-level selection meets social choice theory. *Biology & Philosophy*, 24(5):561–584.

Okasha, S. (2015). The relation between kin and multilevel selection: An approach using causal graphs. *The British Journal for the Philosophy of Science*, page axu047.

Olson, M. (1965). *The Logic of Collective Action: Public goods and the theory of groups*. Harvard University Press.

O'Riain, M., Jarvis, J., Alexander, R., Buffenstein, R., and Peeters, C. (2000). Morphological castes in a vertebrate. *Proceedings of the National Academy of Sciences*, 97(24):13194–13197.

O'Riain, M. J., Jarvis, J. U., and Faulkes, C. G. (1996). A dispersive morph in the naked mole-rat. *Nature*, 380(6575):619–621.

Oyama, S. (1992). Ontogeny and phylogeny; a case of metarecapitulation? In Griffiths, P., editor, *Trees of Life*. Springer.

Oyama, S., Taylor, P., Fogel, A., Lickliter, R., Sterelny, K., Smith, K. C., and van der Weele, C. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*. Duke University Press.

Oyama, S. E., Griffiths, P. E., and Gray, R. D. (2001). *Cycles of contingency: developmental systems and evolution.* MIT Press.

Pacheco, J. M., Traulsen, A., and Nowak, M. A. (2006a). Active linking in evolutionary games. *Journal of Theoretical Biology*, 243(3):437–443.

Pacheco, J. M., Traulsen, A., and Nowak, M. A. (2006b). Coevolution of strategy and structure in complex networks with dynamical linking. *Physical Review Letters*, 97(25):258103.

Paul, D. (1992). Fitness: Historical perspectives. In Keller, E. F. and Lloyd, E. A., editors, *Keywords in Evolutionary Biology*, pages 112–114. Harvard University Press.

Paulsson, J. (2002). Multileveled selection on plasmid replication. *Genetics*, 161(4):1373–1384.

Pepper, J. W. (2000). Relatedness in trait group models of social evolution. *Journal of Theoretical Biology*, 206(3):355–68.

Pepper, J. W. and Herron, M. D. (2008). Does biology need an organism concept? *Biological reviews of the Cambridge Philosophical Society*, 83(4):621–7.

Pigliucci, M. (2007). Do we need an extended evolutionary synthesis? *Evolution*, 61(12):2743–2749.

Platt, T. G. and Bever, J. D. (2009). Kin competition and the evolution of cooperation. *Trends in Ecology & Evolution*, 24(7):370–7.

Popper, K. (1945). *The open society and its enemies*. Routledge.

Powers, S. T. (2010). *Social Niche Construction : Evolutionary Explanations for Cooperative Group Formation*. PhD thesis, University of Southampton.

Powers, S. T. and Lehmann, L. (2013). The co-evolution of social institutions, demography, and large-scale human cooperation. *Ecology letters*, 16(11):1356–1364.

Powers, S. T., Penn, A. S., and Watson, R. A. (2011). The Concurrent Evolution of Cooperation and the Population Structures That Support It. *Evolution*, 65(6):1527–1543.

Pradeu, T. (2013). Immunity and the emergence of individuality. In Bouchard, F. and Huneman, P., editors, *From Groups to Individuals: Evolution and Emerging Individuality*. MIT Press.

Price, G. R. (1970). Selection and covariance. *Nature*, 227:520–21.

Queller, D. (1997). Cooperators Since Life Began. *Quarterly Review of Biology*, 72(2):184–188.

Queller, D. (2000). Relatedness and the fraternal major transitions. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, 355(1403):1647–55.

Queller, D. C. and Strassmann, J. E. (2009). Beyond society: the evolution of organismality. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 364(1533):3143–55.

Rainey, P. and Kerr, B. (2011). Conflicts among levels of selection as fuel for the evolution of individuality. In Calcott, B. and Sterelny, K., editors, *The major transitions in evolution revisited*. MIT Press.

Rainey, P. B. and Kerr, B. (2010). Cheats as first propagules: A new hypothesis for the evolution of individuality during the transition from single cells to multicellularity. *Bioessays*, 32(10):872–880.

Rainey, P. B. and Rainey, K. (2003). Evolution of cooperation and conflict in experimental bacterial populations. *Nature*, 425(6953):72–74.

Rankin, D., Rocha, E., and Brown, S. (2011). What traits are carried on mobile genetic elements, and why? *Heredity*, 106(1):1–10.

Ratnieks, F. L. (1988). Reproductive harmony via mutual policing by workers in eusocial hymenoptera. *American Naturalist*, 132:217–236.

Reeve, H. K. and Sherman, P. W. (1993). Adaptation and the goals of evolutionary research. *Quarterly Review of Biology*, pages 1–32.

Reeve, H. K., Westneat, D. F., Noon, W. A., Sherman, P. W., and Aquadro, C. F. (1990). Dna fingerprinting reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. *Proceedings of the National Academy of Sciences*, 87(7):2496–2500.

Rice, S. H. (2004). *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sinauer Associates Sunderland.

Ridley, M. (2003). *Evolution*. Blackwell.

Rosas, A. (2010). Beyond inclusive fitness? On a simple and general explanation for the evolution of altruism. *Philosophy & Theory in Biology*, 2.

Rosenberg, A. and Bouchard, F. (2010). Fitness. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, fall 2010 edition.

Rosenberg, A. and McShea, D. W. (2007). *Philosophy of Biology: a contemporary introduction*. Routledge.

Roughgarden, J. (2009). *The genial gene: Deconstructing Darwinian selfishness*. University of California Press.

Roze, D. and Michod, R. E. (2001). Mutation, multilevel selection, and the evolution of propagule size during the origin of multicellularity. *The American naturalist*, 158(6):638–54.

Ruiz-Mirazo, K., Etxeberria, A., Moreno, A., and Ibáñez, J. (2000). Organisms and their place in biology. *Theory in Biosciences*, 119(3-4):209–233.

Russell, B. (1927). *An Outline of Philosophy*. Routledge.

Sachs, J. L., Mueller, U. G., Wilcox, T. P., and Bull, J. J. (2004). The evolution of cooperation. *The Quarterly Review of Biology*, 79(2):135–160.

Santelices, B. (1999). How many kinds of individual are there? *Trends in ecology & evolution*, 14(4):152–155.

Santos, F. C., Pacheco, J. M., and Lenaerts, T. (2006a). Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology*, 2(10):e140.

Santos, F. C., Pacheco, J. M., and Lenaerts, T. (2006b). Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proceedings of the National Academy of Sciences*, 103(9):3490–4.

Schnable, P. S. and Wise, R. P. (1998). The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends in Plant Science*, 3(5):175–180.

Sherman, P. W., Jarvis, J. U., and Alexander, R. D., editors (1991). *The Biology of the Naked Mole-Rat*. Princeton University Press.

Sherman, P. W., Jarvis, J. U., and Braude, S. H. (1992). Naked mole rats. *Scientific American*, 267:72–78.

Silver, M. and Di Paolo, E. (2006). Spatial effects favour the evolution of niche construction. *Theoretical Population Biology*, 70(4):387–400.

Simon, H. A. (1981). *The Sciences of the Artificial*. MIT Press.

Simpson, C. (2011a). How many levels are there? how insights from evolutionary transitions in individuality help measure the hierarchical complexity of life. In *The Major Transitions Revisited*, pages 199–225. MIT Press.

Simpson, C. (2011b). The evolutionary history of division of labour. *Proceedings of the Royal Society B: Biological Sciences*.

Simpson, G. G. (1957). *Life: An Introduction to Biology*. New York, NY: Harcourt Brace Jovanovich.

Skyrms, B. (1994). Darwin meets the logic of decision: Correlation in evolutionary game theory. *Philosophy of Science*, 61:503–528.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Sober, E. (1984). *The Nature of Selection*. MIT Press.

Sober, E. (1992). The evolution of altruism: Correlation, cost, and benefit. *Biology and Philosophy*, 7(2):177–187.

Sober, E. (1993). *Philosophy of Biology*. Oxford University Press.

Sober, E. (2001). The two faces of fitness. In Singh, Krimbas, Paul, and Beatty, editors, *Thinking about evolution: Historical, philosophical, and political perspectives*, volume 2, pages 309–321. Cambridge University Press.

Sober, E. and Wilson, D. (1998). *Unto Others*. Harvard University Press.

Sober, E. and Wilson, D. S. (2011). Adaptation and natural selection revisited. *Journal of Evolutionary Biology*, 24(2):462–468.

Sotos, A., Vanhoof, S., Van Den Noortgate, W., and Onghena, P. (2009). The transitivity misconception of pearson's correlation coefficient. *Statistics Education Research Journal*, 8(2):33–55.

Spencer, H. (1864). *Principles of Biology*. Williams and Norgate.

Sterelny, K. and Griffiths, P. (1999). *Sex and Death*. University of Chicago Press.

Taylor, C. and Nowak, M. A. (2007). Transforming the dilemma. *Evolution*, 61(10):2281–2292.

Thompson, J. (2005). *The Geographic Mosaic of Coevolution*. University of Chicago Press.

Traulsen, A. and Nowak, M. A. (2006). Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences*, 103(29):10952–10955.

Triandis, H. (1995). *Individualism and Collectivism*. Westview Press, Colorado.

Trivers, R. L. (1971). The Evolution of Reciprocal Altruism. *Quarterly Review of Biology*, 46(1):35–57.

Tudge, S., Watson, R., and Brede, M. (2016). Game Theoretic Treatments for the Differentiation of Functional Roles in the Transition to Multicellularity. *Journal of Theoretical Biology*.

Umen, J. G. (2014). Green algae and the origins of multicellularity in the plant kingdom. *Cold Spring Harbor perspectives in biology*, 6(11):a016170.

Van Dyken, J. D. and Wade, M. J. (2012). Origins of altruism diversity II: Runaway coevolution of altruistic strategies via 'reciprocal niche construction'. *Evolution*, 66(8):2498–2513.

Van Valen, L. (1976). Energy and evolution. *Evolutionary Theory*, 1(1):l9–229.

Van Valen, L. (1980). Evolution as a zero-sum game for energy. *Evolutionary Theory*, 4:289–300.

Van Valen, L. (2003). Fitness - scalar or vector? *Evolutionary Theory*, 12:155–157.

Van Valen, L. and Sloan, R. E. (1966). The extinction of the multituberculates. *Systematic Biology*, 15(4):261–278.

Van Valen, L. M. (1989). Three paradigms of evolution. *Evolutionary theory*, 9(2).

Velicer, G. J. and Vos, M. (2009). Sociobiology of the myxobacteria. *Annual review of microbiology*, 63:599–623.

von Mises, L. (1949). *Human Action: A Treatise on Economics*. Yale University Press.

Vrba, E. S. (1984). What is species selection? *Systematic Zoology*, 33:318–328.

Vrba, E. S. (1989). Levels of selection and sorting with special reference to the species level. In Harvey, P. and Partridge, L., editors, *Oxford Surveys in Evolutionary Biology*, volume 6, pages 111–168. Oxford University Press.

Weibull, J. (1995). *Evolutionary Game Theory*. MIT Press.

Weismann, A. (1889). *Essays Upon Heredity and Kindred Biological Problems*. Clarendon Press, Oxford.

Weismann, A. (1903). *The Evolution Theory*. London: Edward Arnold.

Wenseleers, T., Hart, A. G., and Ratnieks, F. L. (2004). When resistance is useless: policing and the evolution of reproductive acquiescence in insect societies. *The American Naturalist*, 164(6):E154–E167.

Wernegreen, J. J. and Moran, N. A. (2001). Vertical transmission of biosynthetic plasmids in aphid endosymbionts (buchnera). *Journal of Bacteriology*, 183(2):785–790.

West, S., Fisher, R., Gardner, A., and Kiers, E. (2015). Major evolutionary transitions in individuality. *Proceedings of the National Academy of Sciences*.

West, S. A., Diggle, S. P., Buckling, A., Gardner, A., and Griffin, A. S. (2007a). The Social Lives of Microbes. *Annual Review of Ecology, Evolution, and Systematics*, 38(1):53–77.

West, S. A., Griffin, A. S., and Gardner, A. (2007b). Evolutionary Explanations for Cooperation. *Current Biology*, 17:R661–R672.

West, S. A. and Kiers, E. T. (2009). Evolution: what is an organism? *Current Biology*, 19(23):R1080–R1082.

Williams, G. (1966). *Adaptation and Natural Selection*. Princeton University Press.

Williams, G. and Williams, D. (1957). Natural Selection of Individually Harmful Social Adaptations Among Sibs With Special Reference to Social Insects. *Evolution*, 11(1):32–39.

Williams, G. C. (1992). *Natural Selection: Domains, Levels and Challenges*. Oxford University Press.

Wilson, D. S. and Sober, E. (1989). Reviving the superorganism. *Journal of Theoretical Biology*, 136(3):337–356.

Wilson, D. S. and Wilson, E. O. (2007). Rethinking the theoretical foundation of sociobiology. *The Quarterly Review of Biology*, 82(4):327–348.

Wilson, E. O. and Hölldobler, B. (2005). Eusociality: origin and consequences. *Proceedings of the National Academy of Sciences of the United States of America*, 102(38):13367–13371.

Wilson, D.S. (1975). A theory of group selection. *Proceedings of the National Academy of Sciences*, 72(1):143–146.

Wilson, E.O. (1975). *Sociobiology: The new synthesis*. Harvard University Press.

Winston, J. E. and Jackson, J. B. (1984). Ecology of cryptic coral reef communities. iv. community development and life histories of encrusting cheilostome bryozoa. *Journal of experimental marine biology and ecology*, 76(1):1–21.

Wolstenholme, D. R. (1992). Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, 141:173–216.

Worden, L. and Levin, S. A. (2007). Evolutionary escape from the prisoner's dilemma. *Journal of Theoretical Biology*, 245(3):411–422.

Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16(2):97.

Wynne-Edwards, V. C. (1962). *Animal dispersion in relation to social behaviour*. Oliver and Boyd, London.