Deconstructing evolutionary game theory: coevolution of social behaviors with their evolutionary setting

Erol Akçay*

Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104

September 6, 2019

* e-mail: eakcay@sas.upenn.edu; ORCID: 0000-0001-8149-7124

1

Abstract

Evolution of social behaviors is one of the most fascinating and active fields of evolutionary biology. During the past half century, social evolution theory developed into a mature field with powerful tools to understand the dynamics of social traits such as cooperation under a wide range of conditions. In this paper, I argue that the next stage in the development of social evolution theory should consider the evolution of the setting in which social behaviors evolve. To that end, I propose a conceptual map of the components that make up the evolutionary setting of social behaviors, review existing work that considers the evolution of each component, and discuss potential future directions. The theoretical work reviewed here illustrates how unexpected dynamics can happen when the setting of social evolution itself is evolving, such as cooperation sometimes being self-limiting. I argue that a theory of how the setting of social evolution itself evolves will lead to a deeper understanding of when cooperation and other social behaviors evolve and diversify.

Introduction

All of life is social (Frank, 2007). No organism exists in isolation of others, and all evolve in consequence of their interactions with other organisms. Reflecting this fact, social evolution theory has become an integral part of evolutionary biology, even though it is a relative newcomer, having arisen in its current form decades after the modern synthesis (Hamilton, 1963, 1964; Wilson, 1975). This theory has been an undisputable success: it opened up countless new questions about biological interactions, and successfully answered a good many of them. There have been many controversies along the way, but social evolution theory has reached a mature state with a well-understood core firmly grounded in population genetics (Van Cleve, 2015). This core theory is being applied to problems across all domains of life, and continues to yield new biological questions. My goal in this paper to start from this hard-earned position of success, and to consider what new questions might social evolution theory address going forward.

The major question that motivated Hamilton and many others after him was: how can natural selection favor cooperation, i.e. costly behaviors that benefit others? This problem is as old as evolutionary theory itself, but in one sense we can consider it solved, at least in theory: we know a lot about under what conditions natural selection favors cooperative behaviors (e.g., see reviews by Lehmann and Keller, 2006*b*; West et al., 2007; Nowak, 2006). Two of the most important of these conditions include high relatedness between partners and long-term interactions that allow partners reward each other for cooperation or punish non-cooperation. Moreover, we also know that in many cases in nature, these conditions are in fact met: e.g., cooperative courtship in wild turkeys is a consequence of highly related males pairing up (Krakauer, 2005). Likewise, rhizobium bacteria inhabiting legume root nodules fix nitrogen for their hosts because their hosts preferentially rewards nodules that do (Kiers et al., 2003; Simms et al., 2006). To be sure, the details vary in interesting ways, and many more specific cases remain to be worked out, but

3

we have a good understanding of the general conditions needed to support cooperation and a robust theoretical toolkit to tackle these specific cases. Simply put, the evolution of cooperation is not the puzzle it used to be.

In fact, with so many different mechanisms that can favor cooperation, one might even be puzzled as to why cooperation hasn't evolved everywhere. A trivial answer to this question is that cooperation has not evolved everywhere because conditions that select for cooperation (even if they are numerous) are not found everywhere. In principle, this is an empirical question that can be answered by measuring the relevant variables (relatedness, number of repeated interactions, etc.) and finding if the conditions for cooperation are met or not. Yet, there is also the theoretical question of when such conditions themselves are selected for, which is what I am concerned with in this paper. To unpack this question, we need to consider the setting in which social evolution takes place, akin to what G.E. Hutchinson called "the ecological theater" in which evolution takes place. This theater is made up of various components that themselves are evolving under natural selection. In Figure 1, I give a breakdown of the five components that organize any interaction and provide the setting for evolution. These components are: the game (or payoff structure), proximate mechanisms of behavior, interaction network (including population subdivision), inheritance mechanisms, and the life history of the organisms involved in the interaction (see caption of Figure 1 for more details). This breakdown is not the only logically possible one, but it represents a useful organization of the aspects of interactions that we know to be important in determining the evolution of social traits.

Social evolution theory conventionally takes most or all of the components described in Figure 1 as exogenously fixed, e.g., makes prior assumptions for what the payoff function is, or how the interaction network is structured, and asks what kind of social behavior evolves under a given set of conditions. In this paper, I argue that we should pay more attention to the fact

4

that these components themselves are traits that co-evolve with the behaviors they influence. These coevolutionary dynamics can drive social evolution in unexpected directions, as some of the results I review below illustrate. Understanding how the components that organize social and ecological interactions evolve will help us understand how the diversity of interactions with different kinds of behaviors and social and ecological outcomes comes about.

In what follows, I review recent work on each of the five components that addresses the question of how they coevolve with a social behavior such as cooperation, synthesize these findings, and point to open questions.

The game

The game, i.e., the set of potential behaviors or phenotypes, and their material outcomes or payoffs is where it all starts. The material payoffs to a given phenotype is the ultimate driver of social selection. Most models of the evolution of cooperation are based on a version of the Prisoner's Dilemma, exemplified by the payoff matrix in eq. (1), where the first number is the payoff to the row player, and the second the payoff to the column player:

$$
\begin{array}{ccc}
 & \text{Cooperate} & \text{Defect} \\
\text{Cooperate} & 5,5 & 1,7 \\
\text{Defect} & 7,1 & 2,2
\end{array}
\quad \cdot \tag{1}
$$

The focus on the Prisoner's Dilemma makes sense as a theory building strategy: in this game, any unilateral change in behavior increases one player's payoff while decreasing the other's.

Figure 1: [Figure 1 goes roughly here]

5

In this sense, the Prisoner's Dilemma embodies the purest form of conflict and a worst case scenario for cooperation. This structure also follows from a simple model where a behavior or trait has a net cost to its bearer, but benefits the partner. Yet, what embodies theoretical intuition is not necessarily the most prevalent situation in nature and several researchers have questioned whether games being played in nature are mostly Prisoner's Dilemmas (Noe, 1990, 2006). This is both an empirical and a theoretical question; I will focus on the latter aspect. The payoff structure of the game (i.e., the entries or even the number of rows and the columns of the payoff matrix) can itself evolve through natural selection acting on individual traits that add or remove strategies, or change payoffs from different strategies. The question then becomes how such traits evolve, and what the resulting evolutionarily stable games, if any, are.

Models of evolving games

Peck and Feldman (1986) developed one of the first models that dealt with a game that is changing with individual behaviors. They modeled a population of individuals that engage in pairwise interactions and can behave in a self- or group-directed way. The payoff from the group-directed behavior is fixed and equally shared by the two individuals while the selfish behavior returned a payoff that is dependent on the frequency of selfish behaviors. This assumption introduces an extra layer of frequency dependence due to direct dependence of the payoffs on the frequency of behaviors in the populations, whereas more commonly the payoffs from each pairing are fixed, but frequency dependence might arise due to changing probabilities of different pairings (Lehmann and Keller, 2006*a*). Peck and Feldman (1986) show that if the past cooperation increases the relative rewards from playing the cooperative strategy now, cooperative behaviors can invade even in panmictic populations. This paper is noteworthy for treating for the first time the reciprocal feedback between social behaviors and the game they evolve under. A more recent

6

model by Hilbe et al. (2018) returned to this theme by modeling a general setting where the game played in a repeated interaction changes stochastically as a function of the strategies played in the last round. Similar to Peck and Feldman, they find that cooperation can evolve if cooperative behaviors tend to lead to more profitable games being played down the line, even if each game by itself could not support cooperation.

A different approach to the evolution of the game was proposed by Worden and Levin (2007), who modeled players that innovate new variations of existing strategies with slightly different payoffs. These new strategies can be taken up by learning or discarded according to the payoffs they give. Worden and Levin show that this gradual evolution of payoff matrix through variations in the strategies available tends to lead away from a Prisoner's Dilemma game, and tend to reduce the conflict between the players. This happens due to the fact that Worden and Levin (2007) assume the payoff increments to each player are drawn independently, so new strategies face no trade-offs between benefiting one or the other player. In this way, the conflict embedded in the Prisoner's Dilemma erodes away as the game explores other, conflict-free strategies. A related model by Huang et al. (2012) considers a similar model to Worden and Levin, but allows old strategies to go extinct (whereas Worden and Levin keep them at low frequencies). They find that populations stabilize at an intermediate diversity of strategies and the pairwise interactions between strategies most frequently consitute Hawk-Dove games, rather than Prisoner's Dilemmas.

The models of Worden and Levin (2007) and Huang et al. (2012) share the feature that the introduction of the mutant does not change the behavior of others immediately. That means a new strategy can only invade if it has a higher payoffs against the opponent strategies in the equilibrium already being played. In Akçay and Roughgarden (2011), we changed this assumption by allowing partners to respond to the new payoff structured that result from the introduction

7

of new strategies, such that new strategies can invade by incentivizing partners to behave differently. For example, in a territorial conflict between two animals, an incentive might be less strict policing of established territory boundaries, so a partner yielding territory could still forage in it. This would reduce the cost of yielding territory and can induce the partner to yield. Providing such incentives are likely to be costly to the focal individual, however. Consequently, we find that in a diploid population and with 2-strategy game, incentives might be able to invade but they cannot fix. This is because to be able to invade, a heterozygous mutant has to provide enough incentive to change partner behavior. But once a partner starts cooperating, they cooperate fully, and therefore no further incentives that a diploid can provide at higher cost can elicit more benefit from them. The result is a population with a variety of games being played depending on the genotypes being matched, and cooperation coexists with defection, despite on average having different payoffs.

More recent work by Stewart and Plotkin (2014) considered the co-evolution of the payoff matrix with conditional strategies in iterated prisoner's dilemma games. It is well-known that cooperation can be sustained via various conditional strategies (such as "Tit-for-Tat") in such a setting if the value of future cooperation exceeds the temptation to defect now (Axelrod and Hamilton, 1981). Stewart and Plotkin (2014) asked what happens when individuals can also evolve their investments into cooperation (i.e., the costs they pay and the benefit they provide). They showed that such coevolution sets up a surprising dynamic: when cooperation is prevalent, selection favors higher cooperative investments because both parties enjoy increased benefits of cooperation. But this increased investment comes with increased costs, and therefore increases the temptation to defect which in turn shrinks the set of conditional strategies that can sustain cooperation. Eventually, the population "falls out" of this set, and cooperation collapses. In contrast to the models above, Stewart and Plotkin's (2014) results show that payoff evolution

8

does not necessarily act to resolve conflicts, and may in fact exacerbate them if the underlying conflict is "masked" through the behavioral mechanisms such as reciprocity.

Evolving games through resource feedbacks

In many ecological and social interactions, payoffs depend on the environment, which may change by strategies played in the population. The classical example is a common pool resource that is being exploited by a population playing a variety of harvesting strategies such as low, sustainable harvest effort vs. high, unsustainable effort. (Ostrom, 1990; Sethi and Somanathan, 1996; Tavoni et al., 2012; Tilman et al., 2017). An equivalent biological example is viruses with different virulence levels exploiting host cells (Kerr et al., 2006; Rankin et al., 2007; Lion and Boots, 2010). Alternatively, the resource might be produced at different rates by individuals of differeny phenotypes, as is common in microbial communities (Estrela et al., 2018). In the simplest cases, the payoffs to each harvesting or production strategy only depends on the resource stock and not directly on the frequency of other strategies. In a resource harvesting scenario, this means that if the cost of harvesting is greater the benefits from the harvest, overharvesting will always be stable. Sethi and Somanathan (1996) observed that sustainable harvesters can avert this outcome by punishing overharvesters. Such punishment introduces a direct strategic interaction between the strategies, where the payoffs to individuals depend both on the resource stock and the frequency of strategies. Sethi and Somanathan (1996) show that the resultant coevolutionary dynamics of the resources and strategies can have stable equilibria where cooperators and a high standing resource level are maintained, but there are also stable equilibria where defectors are fixed, and in some cases, the resource goes extinct.

Tavoni et al. (2012) considered a slightly different punishment mechanism where defectors are excluded from community benefits produced by cooperators. Such sanctioning introduces

9

positive frequency dependence in the strategy dynamics, and means that cooperators have to somehow get above a threshold frequency for cooperation to be stably maintained (sometimes in mixed equilibrium with defection). Although ostracism can maintain cooperation, later work by Tilman et al. (2017) shows that cooperative norms that maximize the aggregate welfare (the maximum sustainable profit) are more vulnerable to invasion by defectors, and a more robust cooperative norm needs to overexploit the resource somewhat to reduce the temptation to would-be defectors.

In the papers discussed above, the co-evolution of the resource and strategies generally results in a stable outcome. A recent paper by Weitz et al. (2016) demonstrates that this need not be the case, and persistent oscillations in the strategies and resource levels can be obtained. In their model, the game at any instant is given by an average of two different games, weighted by the environmental state, whose dynamics can be slower or faster than strategy dynamics. Weitz et al. analyze this model exhaustively, documenting several different dynamical regimes. One of the most intruiging ones is one with oscillatory dynamics, which happens when the strategies that are successful in one game cause the environmental state to move in the direction of the other game. Interestingly, Weitz et al. show that the presence of oscillations are independent of the relative time-scales of the environment and strategy dynamics. This happens because in Weitz et al.'s (2016) model, for almost all strategy frequencies, the environment tends to move to one of the boundary states regardless of the strategies being played. The dynamics therefore travel between the corner equilibria of strategy-resource pairings in a heteroclinic cycle.

In a related model, Rand et al. (2017) look at the co-evolution of deliberative and intuitive behavioral strategies in a setting where the costs and benefits from each is co-evolving with their frequency. They also demonstrate oscillations between the strategies, under the assumption that each strategy becomes more costly as its adoption increases in the population. In their model,

10

however, cycling happens only when the costs change sufficiently slowly relative to strategy dynamics, in contrast to the results of Weitz et al. (2016). This is because the dynamics of the costs always tend to an internal equilibrium. But if costs are too slow to catch up with the popularity of strategies, the latter alternatively overshoot and undershoot what would have been their equilibrium abundance. Tilman et al. (2019) recently generalized this approach to include the intrinsic growth or decay dynamics of the environmental resources. Their analysis includes as special cases many of the cases discussed above, and reveals that the dynamical regimes are determined by four parameters that describe the incentives faced by individuals at the corners of the environment-strategy space plus the relative speed of environmental dynamics.

Evolution of incentives under private information and mechanism design

The papers above all treat scenarios with symmetric agents, where no individual can affect the game played more than the other, and there is (effectively) perfect information about the payoffs. If one party controlled the incentives, it would not be surprising to find the game changing in that party's favor. Imperfect and assymetric information, however, can change this picture dramatically, as the party who controls the incentives may not be able to compute what the optimal incentives are from their own perspective. This scenario underlies a large literature on signaling of need by offspring to their parents (Godfray et al., 1991; Godfray, 1995). A crucial problem in this case is how to elicit the information that only the signaler has when the signaler by default has incentives to misrepresent (e.g., the offspring appearing more needy than it really is). The approach in these models is to find a response to signals for the receivers which makes it such that signaler's strategy is a a strictly increasing function of their need. At such "separating equilibria," the receiver can then behave optimally according to its own payoff given the revealed need of the signaler.

11

Nöldeke and Samuelson (1999) showed that the unique separating equilibrium in the parent offspring signaling model of Godfray et al. (1991) results from the parent "charging" the offspring the difference between the marginal fitness value of additional provisioning to the parent and offspring, so that the offspring's optimal demand matches with the parent's. This is the only signaling equilibrium in this model, because the parent is imposing the costs and therefore they can only depend on the signal value and not on the real state of the offspring. Further, the offspring cannot respond to different incentives by the parent, because both the parental cost schedule and offspring behavior are evolving at the same time-scale. But although this signaling equilibrium allows the parent to extract information from the offspring, the costs at the equilibrium may be so high that the parent is better off without any signaling (Rodríguez-Gironés et al., 1996; Johnstone, 1999) and the signaling equilibrium may not be robust to finite perturbations (Rodríguez-Gironés et al., 1998).

One potential solution to this problem is to consider the fact that signaling involves some learning: parents likely evolve a reaction norm that describes how much they will feed the offspring for a given signaling level, and the offspring can adjust its signaling level to maximize its net growth (Akçay, 2012). In such a scenario, parents can evolve to impose lower costs for the same amount of provisioning. This would lead parents to do more provisioning than would maximize their fitness if they had perfect information about offspring need, but they can still have hifher fitness than the signaling equilibrium of Godfray et al. (1991). The offspring's ability to extract more resources than would have been optimal for the parent is called an "information rent" in economics.

More generally, costly signaling theory is closely linked to a branch of economics called mechanism design (Vickrey, 1961; Myerson, 1979). Mechanism design asks the question of how to design a game to elicit a desired outcome from optimally behaving agents. A central theorem

12

of mechanism design, called the "revelation principle" (Myerson, 1979) allows one to represent the equilibria of any game, however complicated, with a given information structure (i.e., who knows what) with one from a class of simple games. In these games, called "direct mechanisms," individuals simply reveal their information to a central arbiter and receibe their payoffs. The only constraint is that the function mapping information revealed to payoffs should make it optimal for the individuals to "reveal" their private information honestly, a condition called "incentive compatibility." Incentive compatible direct mechanisms allow one to characterize the entire range of possible outcomes in a given information environment. Crucially, one can say what outcomes are impossible in any game, no matter how complex or clever.

Myerson and Satterthwaite (1983) famously applied the revelation theorem to the case of a buyer and seller bargaining over the sale of an item. They showed that if the player's valuation of the item is only known to them, there is no game (bargaining procedure) that will make sure the trade will take whenever it is mutually beneficial. In other words, some mutually beneficial exchanges are completely precluded because of the problem of private information. Akçay et al. (2012) showed that the setting considered by Myerson and Satterthwaite is closely related to the reproductive transactions theory of cooperative breeding (Vehrencamp, 1983). Reproductive transactions theory has numerous models with particular assumptions that can determine reproductive transactions in such a setting. But none had considered the possibility that individual might be privately informed about their expected success in breeeding alone. Akçay et al. (2012) show that the Myerson and Satterthwaite result applies to reproductive transactions as well: regardless of the details of the game structure, private information may preclude individuals from cooperating even when it is mutually beneficial to do so.

13

Future directions for evolving games

The above subsections provide a review of models that explicitly address how the payoffs from a game evolve. So far, research on this question has proceeded in somewhat disjoint literature, partly reflecting the fact that there are numerous ways in which the game can evolve (e.g., through resource feedbacks or through evolving new strategies), each raising different theoretical issues. Moving forward will entail exploring these different ways in which the game can evolve and integrating results across settings.

One common theme emerging from the papers discussed above that game evolution can resolve some of the conflicts of interests embodied in the Prisoner's Dilemma game, and hence create more favorable conditions for cooperation. However, the resolution may not be complete because either the modifications to the payoff matrix themselves are constrained by underlying trade-offs (as in Akçay and Roughgarden, 2011), or previous evolution has locked players into a Hawk-Dove game (as in Worden and Levin, 2007; Huang et al., 2012). At the end of the day, the world has finite resources, which will impose some unbreakable trade-offs between players at some level. The Prisoner's Dilemma is supposed to distill such conflicts, but the above results show that social systems can evolve to deal with this fundamental conflict in different ways. At the same time, payoff conflicts can be resolved through various mechanisms (such as conditional behaviors that cast "the shadow of the future" over current interactions), but by masking the underlying conflict such mechanisms might lead to increased conflict in the long-run (Stewart and Plotkin, 2014). These results illustrate that the evolutionary dynamics of payoff conflict and alignment between agents can take unexpected turns.

Economists and political scientists working on the evolution of institutions have grappled with resolving conflicts of interests for a long time (North, 1991; Bowles et al., 2003). Integrating their tools with evolutionary theory is likely to be fruitful to understand how social systems

14

evolve to manage conflicts of interests (Akçay et al., 2010, 2012). A major question for such a theory is how much conflict and in which form (e.g., a game of pure conflict such as Prisoner's Dilemma vs. coordination games with partial conflict like Hawk-Dove) will persist for a given shape of fundamental trade-offs that underlie the interaction. Other aspects of the game, such as the roles of players, the specificity of such roles (Bshary and Bronstein, 2004; Rodrigues and Kokko, 2016), or the order in which players take actions (Pen and Taylor, 2005; McNamara et al., 2006) are also evolvable, but how selection act on these remains largely unexplored. Finally, most of the work above has also taken place in discrete-action, two player games; a broader diversity of models including continuous games and N-player interactions likely will reveal more surprising dynamics.

Proximate mechanisms

The second component of Figure 1 we will take up is the proximate mechanisms of behavior. In this context, proximate mechanism refers to any biological mechanism that produces a behavior or phenotype within the lifetime of an organism or the duration of an interaction. Although it is universally admitted that proximate (mechanistic) and ultimate (evolutionary) causation of traits need to be integrated (Tinbergen, 1963), until recently proximate mechanisms mostly took a back seat in the development of social evolution theory (with some notable exceptions such as reciprocal altruism (Trivers, 1971; Axelrod and Hamilton, 1981) and developmental plasticity (West-Eberhard, 1989)). Theory in behavioral ecology proceeded for a long time under the working assumption that natural selection will somehow take care of the proximate mechanism to produce ecologically optimal behaviors (Grafen, 1984). But such a strict dichotomy of proximate and ultimate causes neglects that "ultimate" selection pressures can be determined by "proximate" mechanisms, which suggests reciprocal cycles of causation with no clear hierarchy (Laland

15

et al., 2011, 2013). Behavioral ecology theory has been making progress in moving beyond this dichotomy (Stamps, 1991; Roughgarden et al., 2006; Roughgarden, 2009; McNamara and Houston, 2009; Fawcett et al., 2012; McNamara, 2013). A sizable literature now exists that explicitly models (simple) proximate mechanisms of behavior and the evolution of such proximate mechanisms. In this section, I focus on models of simple proximate mechanisms in continuous social dilemmas that illustrate some of the still open questions.

Evolution of behavioral dynamics in continuous games

To fix the setting, imagine a typical social dilemma where individuals can invest continuously varying amounts into behaviors that are costly for themselves but benefit others. How will individuals decide on how much to invest? A collection of models (starting with Doebeli and Knowlton, 1998; Wahl and Nowak, 1999; McNamara et al., 1999) investigate simple proximate mechanisms where individuals' investments are determined by a linear reaction norm, e.g., one specifying that a focal individual matches the partner contribution at a certain ratio. In these models, individuals start with some set of investments, and keep updating their behavior in response to each other, eventually arriving at a stable pair of actions. The actions at this "behavioral equilibrium" then determine the final payoff, and therefore fitness of the individuals. The behavioral equilibrium in turn is a function of the slope and intercept of the linear reaction norms. This slope and intercept, rather than any particular action, are heritable traits that evolve according to the behavioral equilibria they produce. Thus, these models put evolution at an arm's length from the actual behaviors. The same genotype (describing a reaction norm) can behave differently according to the partners' genotype, making these models special cases of indirect genetic effects (Moore et al., 1997; McGlothlin et al., 2010; Akçay and Van Cleve, 2012). Cooperation can be stable in such games even in the absence of relatedness, since individuals can

16

evolve to behaviorally respond to one another and reduce or increase contributions depending on partners' actions. These behavioral responses counteract the incentives to withhold cooperation and reward partners that cooperate (amounting to reciprocity).

There are two main theoretical findings in this basic setup. The first is that there is a continuum of evolutionarily stable reaction norms that can result in any level of cooperation from zero to maximally beneficial (Taylor and Day, 2004). This is because one can always find a slope-intercept pair that generates a behavioral equilibrium that is made evolutionarily stable by the slope (Akçay and Van Cleve, 2012; Van Cleve and Akçay, 2014). The second finding is that absent constraints on what slope-intercept pairs can invade, no pair is strictly evolutionarily stable, because there is an infinite number of slope-intercept pairs that produce the same behavioral outcome against the resident (Taylor and Day, 2004). Some of these reaction norms, when playing against themselves, will result in higher cooperation than the resident and can invade through a combination of drift and mutant-mutant interactions. This process tends to favor increased cooperation up to the maximum possible total benefit (André and Day, 2007). Interestingly, the multiplicity and neutrality of equilibria disappear when the payoff function fluctuates due to environmental variation (McNamara et al., 1999; Johnstone and Hinde, 2006; André, 2015; Ito et al., 2017). In these cases, the linear reaction norm has to either be best response across a range of behaviors (McNamara et al., 1999; Johnstone and Hinde, 2006), or be optimal in expectation (André, 2015; Ito et al., 2017), which gets rid of the degeneracy.

An alternative approach is to model organisms as goal-oriented agents that try to behaviorally maximize some objective function, representing intrinsic motivations of agents. These motivations are in turn shaped by natural selection (Akçay et al., 2009; Akçay and Van Cleve, 2012; Van Cleve and Akçay, 2014). This approach captures the notion that animals have internal representations of rewards of outcomes and engage in goal-seeking behaviors (McFarland and Sibly,

17

1975; McFarland, 1991). These models show that the shape of the objective function and what aspects of it are evolving crucially determines the evolutionary outcome. For example, Akçay et al. (2009) considered the evolution of other-regarding preferences in two-person games, and show that they cannot evolve in the additive form (i.e., a weighted sum of own and partners pay-off) but can in a multiplicative form (a weighted product of the two payoffs). This is because the latter type produces positive responsiveness (analogous to the slope of the linear reaction norms) at behavioral equilibria with positive investment, which incentivizes the partner to investing in cooperation. In contrast, an additive other-regard function produces a negative responsiveness and therefore does not incentivize cooperation. Akçay and Van Cleve (2012) extended this model to multiplayer interactions between related individuals, and showed that the the responsiveness coefficient plays exactly the same role as relatedness in determining the evolutionary outcome. This framework can also be mapped directly on the formalism of indirect genetic effects (Mc-Glothlin et al., 2010), and thus provides a way to understand the evolution of the indirect genetic effects (Kazancıoğlu et al., 2012).

Future directions for modeling proximate mechanisms

The good news from the theoretical work reviewed so far is that including proximate mechanisms directly into evolutionary models can be done in a tractable way, and has yielded interesting insights in how the proximate basis of social behaviors can evolve. The bad news is that the predicted outcomes are often not unique and depend on details of the proximate mechanisms assumed in the models. Developing a predictive theory of what kind of proximate mechanism will evolve, and what type of genetic variation will be available to natural selection remain the biggest open questions.

One way to address this question is to model the mechanistic basis of how decisions are

18

made, e.g., through modeling dynamics of reinforcement learning (Dridi and Lehmann, 2015; Dridi and Akcay, 2018) or neural networks (Enquist and Ghirlanda, 2005). For example, McNally et al. (2012) studied the evolution of neural networks that play iterated (discrete action) Prisoner's Dilemma and Snowdrift games with each other. They allow these networks to evolve both by changing network weights and adding and removing nodes that compute responses or record the states of other nodes, the latter allowing for the network to have longer memory. They find that selection tends to increase the complexity of the network when the game is near its (single-shot) Nash equilibrium. Although more complex networks do not necesarily lead to more cooperation on average, the variance of cooperation increases. One reason that frequency of different outcomes can change with differences in the complexity of the underlying mechanism is that the mutational distance between strategies are different (van den Berg and Weissing, 2015; Stewart and Plotkin, 2015). How many mutations it takes to go from one kind of conditional strategy to another determines which strategies can compete with each other and therefore affects the stability of each.

These results further highlight that the details of the proximate mechanisms that produce flexible behaviors can affect the evolutionary outcomes. This might worry some theorists, as it means that the space of possible outcomes is not well-constrained. But a more positive interpretation is that modeling proximate mechanisms introduces a rich space of outcomes with structure that we can explore. More importantly, new methods for fine-scale quantification of behavioral interactions (e.g., Klibaite et al., 2017) can be used to constrain our models of proximate mechanisms, although we are not at the point of connecting our evolutionary models to such behavioral data yet. In other contexts (e.g., microbial social evolution and social insects), gene regulatory networks as opposed to neural networks may be the more appropriate mechanistic basis for modeling flexible behaviors (Akçay et al., 2015). This is another context where we

19

have the advantage of increasingly fine-scale look at how gene expression changes with social interactions. To take advantage of these new data from an evolutionary perspective, we need to combine our evolutionary models with better models of proximate mechanisms, which is where a lot of exciting new work remains to be done.

Interaction networks

The third component of social evolution theory in Figure 1 is the interaction network structure, defined very broadly as who interacts with whom. The most studied effector of social evolution, genetic relatedness between individuals, is a consequence of interaction network structure, and is the subject of a vast theoretical literature (Rousset and Billiard, 2000; Rousset, 2004; Tarnita et al., 2009; Lehmann and Rousset, 2010; Van Cleve, 2015). In this section, I review recent research that has started to push into two directions: the co-evolution of the life-history variables that affect population structure with the social traits, and the structure and evolution of fine-scale social structure within social groups.

Evolution of dispersal rates in structured populations

The central role dispersal plays both as a driver of social selection (Hamilton, 1964) and as a trait under social selection itself (Hamilton and May, 1977) has long been recognized. One of the most important theoretical insights after Hamilton was that although limited dispersal can increase relatedness and favor altruism, it also increases local competition with kin which disfavors altruism (Wilson et al., 1992; Taylor, 1992). Considering the coevolution of dispersal and behaviors therefore requires accounting for the long-term fitness consequences on a spatially extended lineage. Perrin and Lehmann (2001) investigated the coevolution of dispersal and cooperation

20

under kin-recognition in patch structured populations. They found that kin recognition can alleviate some of the canceling effect of local competition, and depending on the mechanism of kin recognition, limited dispersal may have a negative or non-monotonic relationship to altruism at the coeovolutionary equilibrium. Le Galliard et al. (2005) took up the same question under a model where empty sites were allowed, which relaxes the negative effect of local competition, and showed that when costs to mobility are high, cooperative behavior and low dispersal can co-evolve and reinforce each other.

Both of the above models use an adaptive dynamics framework where the population is almost always monomorphic in dispersal; i.e., cooperators and defectors have the same dispersal traits. Koella (2000) used an individual-based simulation model of a population living on a lattice, and allowed heterogeneity in dispersal rates. In his model, starting with a low dispersal rate, cooperation readily evolves. However, when the dispersal rates are allowed to co-evolve, defectors evolve high dispersal rate, which facilitates their invasion of cooperative subpopulations, resulting in a polymorphism between cooperators that disperse locally and defectors that disperse long-distance. A similar result is found by Hochberg et al. (2008) who considered a group-structured model where cooperators invest in public goods that promote the growth of the group: in their model, cooperators evolve to stay in groups while defectors evolve to disperse out. Similarly, Smaldino et al. (2013) find in a spatially explicit model that cooperators and defectors can coexist when the former move around less and the latter more.

In these models, cooperators and cheaters are discrete types. Purcell et al. (2012) modeled the coevolution of cooperation and dispersal as continuous traits in patch-structured populations. Using individual-based simulations, they show that a negative correlation between cooperation and dispersal traits can evolve, and is stronger when dispersal is costly. That can lead to the emergence of patches of cooperative individuals that stay together, and eventually increase aver-

age cooperation levels in the group. Interestingly, Purcell et al. find that these dynamics depend crucially on the mode of population regulation: cooperation and low dispersal are harder to evolve if density dependence acts at the patch level (i.e., wipes out entire patches) compared to at the individual level (i.e., individuals from random patches die due to density dependence.) Parvinen (2013) considers a related model with both patch-level density dependence and positive probabilities of entire patches getting wiped out. He studies the conditions for evolutionary branching in the cooperation and dispersal traits that gives rise to two separate types as in the discrete model above. Parvinen shows that this can happen to cooperation when the benefit from cooperation is saturating in the number of cooperators, due to the ecological dynamics of group size. In the resulting polymorphic population, cooperators then evolve lower dispersal, and defectors higher. More recently, Mullon et al. (2018) modeled the co-evolution of the cooperation probability with dispersal. In this setting, the payoff from cooperation is linear in the probability to cooperate, so evolutionary branching does not happen when either cooperation or dispersal are evolving alone. Strikingly, however, evolutionary branching happens under a wide range of conditions when the probability of cooperation and dispersal are both evolving, as mutants that have higher cooperation and lower dispersal and vice versa can both invade an evolutionary attractor point. Interestingly, they also found that recombination between cooperation and dispersal loci will then be selected to decrease so that the two traits will be genetically integrated. This further stabilizes the polymorphism.

Coevolution of network structure with social behaviors

The papers discussed above mostly deal with the physical subdivision of populations. Population structure can also exist at a finer scale: even amongst individuals live in the same patch or group, some pairs will interact more than others. Empirical studies are increasing providing detailed

quantitative picture of such fine scale social structure using tools from social network analysis (Wey et al., 2008; Pinter-Wollman et al., 2013). These social networks are inherently dynamic, as the presence or absence of an interaction results from the decisions organisms make in their lifetimes. Furthermore, these decisions are driven by heritable traits which can coevolve with other social traits. One mechanism that organizes fine-scale social structure is partner choice (Bull and Rice, 1991), which has been considered in several models that regard network dynamics implicitly. McNamara et al. (2008) modeled the coevolution of such choosiness in partners and cooperation levels, and found that low mortality, which reduces the opportunity costs of being choosy, favors partner choice and cooperation. Partner choice can also be affected by institutions that assort cooperators and defectors: Bowles et al. (2003) showed that such institutions can evolve in group structured populations as they promote higher cooperation and group-level performance.

Another group of models looks at network dynamics explicitly and considered how cooperation and network structure coevolve: Ilany and Akçay (2016) proposed a model where newborns make connections by "socially inheriting" their parent's connections and making "random" connections with others. They showed that this process can match important statistics such as degree distribution, clustering coefficients, and modularity of actual networks from four vertebrate species. Additionally, social inheritance automatically generates assortment within the group in (genetically) inherited traits. Cavaliere et al. (2012) considered the evolution of cooperation (providing benefits to connections) in a special case of the social inheritance model that omits random linking. They observed that when cooperation is at high frequency, networks become well-connected, while they were sparsely connected when cooperation was rare. This happens because in cooperative populations, individuals with more connections receive more benefits therefore reproduce more. As their offspring socially inherit their connections, newborns there-

23

fore also tend to be highly connected. Conversely, when cooperation is at low frequency, network connections do not matter for fitness (because defectors provide no benefits), and one recovers a neutral network structure, which for the parameters Cavaliere et al. considered is sparse (Ilany and Akçay, 2016).

In these models, the probabilities of social inheritance and random connection are fixed traits. But in nature, the linking traits themselves are likely to be evolvable traits and may co-evolve with social behaviors, which produces unexpected dynamics (Akçay, 2018). Under these network dynamics, cooperation evolves only under low probability of random linking. However, once cooperation is at high frequency, selection favors making connections indiscriminately because everyone is giving out benefits. Thus, the probability of random linking increases, and eventually stops being able to sustain cooperation. This shows that cooperation selects for conditions that undermines itself. This dynamic can be cut short if social connections are costly, but such costs can negate most or all of the benefits from cooperation (Akçay, 2018).

The overall message from this section is that the coevolutionary dynamics of dispersal and social behaviors can lead to outcomes one might not expect from considering each alone, as is mostly done in social evolution. These coevolutionary dynamics sometimes favor cooperation by preparing conditions conducive to it (e.g. Le Galliard et al., 2005; Purcell et al., 2012); but cooperation can also undermine itself (Akçay, 2018). The emergence of distinct combinations of dispersal strategies and social behaviors (e.g., Koella, 2000; Mullon et al., 2018) can be considered as a social niche (Saltz et al., 2016). Co-evolution of the interaction network with social behaviors seems to promote diversification of social niches, but what determines the diversity of social niches remains an open question (Saltz and Foley, 2011; Montiglio et al., 2013).

24

Demography

The next component from Figure 1 is the demographic structure of populations, which is well-established as a potent modulator of social selection (Lehmann and Rousset, 2010). Prior research on the coevolution of demography with social behaviors focused mainly on the coevolution of social behavior with dispersal (discussed above), population densities or group size, and to a lesser extent, mating systems.

Coevolution of population density and group size with social traits

Social behaviors can change the birth and death rates of populations and therefore affect their ecological dynamics. Most social evolution models deal with density dependence by assuming a fully saturated population of constant size. With local dispersal, this means that local competition between kin cancels some indirect benefits due to increased relatedness, as discussed above. One way to alleviate this problem is to consider "elastic" populations, i.e., ones that can grow or shrink according to the outcome of social dynamics (Wilson, 1987; Lehmann et al., 2006; Platt and Bever, 2009). This phenomenon has been modeled with either group structured populations where the group's carrying capactiy depends on the frequency of cooperation (Wilson, 1987; Lehmann et al., 2006), or in spatially extended populations (such as a lattice) with local reproduction (e.g., Van Baalen and Rand, 1998) and the potential for empty sites. In the latter case, empty sites (e.g., created by background mortality) relax kin competition as the population density co-evolves with the social behavior. These models consider the emergence of correlations between the types of neighboring individuals and the number of empty patches experienced by each type from the eco-evolutionary dynamics of the social traits. A main message from them is that the type of demographic benefit (e.g., fecundity increase vs. survival decrease) from cooperation can change the conditions under which cooperation evolves and its population consequences (Lion

25

and Gandon, 2009; Débarre et al., 2014). One of these consequences is that increased survival or fecundity due to cooperation might constrain further evolution of cooperation due to populations being locally saturated around cooperators, which causes increased kin competition locally.

A similar negative feedback between population density and cooperation can happen in group-structured populations when group size increases with density, as this reduces the average relatedness between individuals in a group (Hauert et al., 2006). Such a negative feedback was demonstrated experimentally by Sanchez and Gore (2013) using budding yeast that secrete invertase into the periplasmic space where it breaks down sucrose into glucose and fructose which then diffuse out to the environment. They showed that cheater strains that do not produce invertase do better at high population densities, while producer strains do well at low densities, but the latter cause population density to increase. Van Dyken and Wade (2012) show populations can escape this negative feedback when multiple social traits co-evolve, where one increases the survival or fecundity of social partners and another increases local resource availability. Survival-increasing cooperation creates higher local population sizes, which leads to selection for increased resource provision. This in turn relaxes local competition and selects for cooperative traits that increase fecundity. This positive feedback can lead to much higher levels of either type of helping in a coevolutionary setting than when evolving in isolation. It is also worth noting that even if population or group sizes are fixed, niche construction can evolve if the benefits from the resource is felt in subsequent generations, as the spreading of benefits in time also alleviates local competition between the producer of the public good and its beneficiaries (Lehmann, 2007, 2008).

In the above models, population density or group size changes as a passive consequence of the social behavior. It is also possible that group size is selected more directly through individual preferences for association. Avilés (2002) modeled a population where groups are formed

through a process of accretion, where each new member joins a group with a probability that is a function of its own and other group members grouping preferences. The coevolution of these preferences with cooperative behaviors that increase group productivity can produce coevolutionary cycles: in populations with low cooperativeness, a negative correlation evolves between cooperation and grouping tendency, which favors cooperators forming smaller groups and keeping the non-cooperators out. As cooperation increases, however, an incentive to join bigger groups emerges. The negative correlation between grouping tendency and cooperation disappears and bigger, more mixed groups form that favor non-cooperators. This is another kind feedback that makes cooperation self-limiting. In a related model, Powers et al. (2011) find a contrasting result: preference for the smallest founding group sizes (single individuals) coevolves with cooperation, and remains stable. The reason is that Powers et al. (2011) consider a "haystack" model where groups grow separately from other groups for a certain number of reproductive cycles of individuals. This sets up selection against larger group sizes in the model, as groups with two founders would run the chance of acquiring a defector and therefore growing less. In contrast, in Avilés (2002), groups disband at the end of every reproductive cycle. Interestingly, Powers et al. find that the initial evolution of small founder sizes might take a long time, because it first needs a positive genetic association between cooperation and smaller group sizes to form. For populations with initially large group size preferences, selection does not favor cooperation, and therefore the necessary associations do not form except through drift.

Coevolution of mating systems with altruism

One of the most interesting recent debates in social evolution in recent years is how mating systems influence the evolution of cooperation between siblings, including eusociality and cooperative breeding (Hughes et al., 2008; Nowak et al., 2010; Cornwallis et al., 2010; Lukas and

27

Clutton-Brock, 2012). Theory predicts that high relatedness due to genetic monogamy or inbreeding favors cooperation within sib groups (Hamilton, 1972; Charlesworth, 1978) but where do these mating systems come from? This question attracted some interest in the late 1980s, but surprisingly little follow-up since. Peck and Feldman (1988) showed in a two locus population genetic model that monogamous mating preferences can coevolve with altruistic behaviors. Likewise, Breden and Wade (1991) showed that preferences for inbreeding (within sibship mating) can also coevolve and mutually reinforce altruistic behaviors. In both cases, a linkage disequilibrium arises between the altruistic allele and the mating preference allele that produces high relatedness, since altruism is selected for within families with high relatedness, and selected against otherwise. In this way, mating preferences that produce high relatedness and altruistic traits can reinforce each other. The question of how mating preferences would coevolve with social behaviors appears ripe to be treated with more recent theoretical tools in multi-locus social evolution theory (Gardner et al., 2007; Roze and Rousset, 2008).

Inheritance mechanisms

The final component of Figure 1 we will take up is the mechanism by which the social behavior or its context is inherited. Most social evolutionary theory explicitly (or sometimes implicitly) assumes genetic inheritance, but it is becoming increasingly clear that other mechanisms of inheritance (e.g., cultural or ecological) can play important roles in evolution (Bonduriansky and Day, 2018). This issue already figured in the discussions above, as in the discussion of social inheritance (Ilany and Akçay, 2016; Akçay, 2018) or niche construction across generations (Lehmann and Balloux, 2007; Lehmann, 2008). There is increasing interest in how different inheritance mechanisms can affect social behaviors, and some recent attempts at general frameworks for modeling the dynamics of traits that are genetically as well as non-genetically inherited. How-

28

ever the question of which inheritance mode will evolve for a social trait has only been addressed sparsely.

One of the most discussed modes of non-genetic inheritance is cultural transmission. Culturally transmitted traits get passed on not just vertically from parents to offspring, but potentially also between other members of a population. They are of obvious importance to human social dynamics, but increasingly recognized as playing a key role in non-human animals' social and ecological adaptations (Whiten, 2019). One of the popular arguments for explaining large-scale human cooperation is cultural group selection (Boyd and Richerson, 1982, 2002, 2005), where cultural processes such as conformist bias reduce within-group variance in phenotypes relative to between-group variance, and sets the stage for spread of group-beneficial cultural traits in the population (but see Lehmann and Feldman, 2008; Van Cleve, 2016, for potential wrinkles in this story). This hypothesis illustrates how a different inheritance mode can change selection on a trait due to generating different population structures (Cavalli-Sforza and Feldman, 1981). Another example of a inheritance mode separate from the genetic inheritance (of the host, in this case) is when social behaviors are influenced by symbionts: Lewin-Epstein et al. (2017) show that a symbiont that is vertically transmitted may evolve to make its carriers behave altruistically towards others if it also can be horizontally transmitted during social interactions, even if such altruism would not have been favored if it was inherited genetically in the host. This happens because once transmitted to a social partner of its previous host, the symbiont benefits from the increased fitness of the recipient as a result of the altruistic act.

In both examples above, depending on the inheritance mode through which a phenotype is transmitted, evolutionary dynamics can go in different directions. Although this fact is well-established now (Danchin et al., 2011), there is relatively little work on how the inheritance mode itself will coevolve with a social (or asocial) trait. One exception is the literature on the

evolution of socially transmitted mating preferences (Servedio and Kirkpatrick, 1996; Tramm and Servedio, 2008; Chaffee et al., 2013). A recent model by Yeh et al. (2018) studied how paternally imprinted mating preferences co-evolve with genetically transmitted preferences and ecological traits under divergent viability selection. They show that sexual imprinting can evolve to generate assortative mating and maintain trait polymorphism, thus enabling local adaptation to different environments to persist. Interestingly, imprinting frequently evolves to be incomplete, with some role of the genetically transmitted preferences persisting, making this a truly co-inherited trait. The balance between the two inheritance systems is to a large extent determined by the degree of divergent selection on the ecological traits and the cost of imprinting.

More generally, answering how the mode of inheritance evolves requires a theoretical framework for modeling the evolutionary dynamics of a trait jointly inherited through genetic and other modes of inheritance. Recent work has started to make progress in this direction (Otto et al., 1995; Bonduriansky and Day, 2009; Helanterä and Uller, 2010; Day and Bonduriansky, 2011; Aguilar and Akçay, 2018), yet more studies like Yeh et al. (2018) are needed that explicitly address how much a given phenotype will evolve to be inherited genetically vs. otherwise. One way the inheritance mode can evolve is by the generation and depletion of variation within each domain (e.g. genetic vs. cultural Aguilar and Akçay, 2018). For instance, if selection depleted genetic variation in a trait or if some mechanism generates culturally heritable variation in that trait, further evolution of the trait would be increasingly dependent on the cultural inheritance system. The joint evolution of inheritance mechanisms and social traits is likely to yield interesting insights.

Conclusion

This review represents an attempt to organize recent efforts that look at the coevolution of social behaviors with the context that determines their evolution. Figure 1 presents a breakdown of the components that make up the evolutionary context, and may coevolve with social behaviors. Like all such divisions, it contains an element of arbitrariness (e.g., dispersal obviously affects both the interaction network and demography components), but I believe it represents a reasonable working map of the different elements all social evolution models need to consider implicitly or explicitly. The main thrust of this paper is that these elements themselves are the result of evolutionary processes, and their evolution can be affected by the social behaviors that they cause the evolution of. This "reciprocal causation" (Laland et al., 2011, 2013) caused by feedbacks between social behaviors and the context they evolve in opens up new questions and has the potential to provide a deeper understanding of how the diversity of social systems in nature and human societies evolves. One of the main conclusions from this review is that the coevolving "theater" of social evolution can cause cooperation (or other traits) to be self-reinforcing or self-limiting, depending on what aspect of the setting is co-evolving and how social behaviors feedback on these. To return to our original question, we would expect cooperation to be stable in the long term when it reinforces the conditions that select for it, and not when it counteracts those conditions.

Finally, although this review is organized to consider the evolution of each component with social behavior separately, in general we would expect multiple components in Figure 1 to coevolve with each other. For instance, the preferences of individuals playing a game can co-evolve with the payoff structure of the game, or the interaction network that the game is played in. Interactions between different factors (such as relatedness, reciprocity, partner choice, and payoff synergy) are known to interact non-additively (Lehmann and Keller, 2006*b*; Van Cleve and Akçay, 2014) and can reinforce or negate each others' effects (Akçay, 2017). It seems likely that

31

considering the coevolution of the components of social evolution will also reveal such interactions.

Acknowledgements

References

Aguilar, E. G., and E. Akçay. 2018. Gene-culture co-inheritance of a behavioral trait. American Naturalist 192.

Akçay, E. 2017. Population structure reduces benefits from partner choice in mutualistic symbiosis. Proceedings of the Royal Society B 284:20162317.

Akçay, E. 2018. Collapse and rescue of cooperation in evolving dynamic networks. Nature Communications 9:2692.

Akçay, E., T. A. Linksvayer, and J. Van Cleve. 2015. Bridging social evolution theory and emerging empirical approaches to social behavior. Current Opinion in Behavioral Sciences 6:59–64.

Akçay, E., A. Meirowitz, K. W. Ramsay, and S. A. Levin. 2012. Evolution of cooperation and skew under imperfect information. Proceedings of the National Academy of Sciences of the USA 109:14936–14941.

Akçay, E., J. Roughgarden, J. D. Fearon, J. A. Ferejohn, and B. Weingast. 2010. Biological institutions: the political science of animal cooperation. SSRN Working Paper 2370952.

Akçay, E. 2012. Incentives in the family II: Behavioral dynamics and the evolution of non-costly signaling. Journal of Theoretical Biology 294:9–18.

Akçay, E., and J. Roughgarden. 2011. The evolution of payoff matrices: providing incentives to cooperate. Proceedings of the Royal Society B 278:2198–206.

Akçay, E., and J. Van Cleve. 2012. Behavioral responses in structured populations pave the way to group optimality. American Naturalist 179:257–269.

Akçay, E., J. Van Cleve, M. W. Feldman, and J. Roughgarden. 2009. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences of the USA 106:19061–6.

André, J.-B. 2015. Contingency in the evolutionary emergence of reciprocal cooperation. American Naturalist 185:303–316.

André, J.-B., and T. Day. 2007. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemma. Journal of Theoretical Biology 247:11–22.

Avilés, L. 2002. Solving the freeloaders paradox: genetic associations and frequency-dependent selection in the evolution of cooperation among nonrelatives. Proceedings of the National Academy of Sciences of the USA 99:14268–14273.

Axelrod, R., and W. D. Hamilton. 1981. The evolution of cooperation. Science 211:1390–1396.

Bonduriansky, R., and T. Day. 2009. Nongenetic inheritance and its evolutionary implications. Annual Review of Ecology, Evolution, and Systematics 40:103–125.

———. 2018. Extended Heredity A New Understanding of Inheritance and Evolution. Princeton University Press, Princeton NJ.

Bowles, S., J.-k. Choi, and A. Hopfensitz. 2003. The co-evolution of individual behaviors and social institutions. Journal of Theoretical Biology 223:135–147.

Boyd, R., and P. J. Richerson. 1982. Cultural transmission and the evolution of cooperative behavior. Human Ecology 10:325–351.

———. 2002. Group beneficial norms can spread rapidly in a structured population. Journal of Theoretical Biology 215:287–96.

———. 2005. The origin and evolution of cultures. Oxford University Press.

Breden, F., and M. J. Wade. 1991. "Runaway" social evolution: reinforcing selection for inbreeding and altruism. Journal of Theoretical Biology 153:323–337.

Bshary, R., and J. L. Bronstein. 2004. Game structures in mutualistic interactions: what can the evidence tell us about the kind of models we need? Advances in the Study of Behavior 34:59–102.

Bull, J. J., and W. R. Rice. 1991. Distinguishing mechanisms for the evolution of co-operation. Journal of Theoretical Biology 149:63–74.

Cavaliere, M., S. Sedwards, C. E. Tarnita, M. A. Nowak, and A. Csikász-Nagy. 2012. Prosperity is associated with instability in dynamical networks. Journal of Theoretical Biology 299:126–138.

Cavalli-Sforza, L. L., and M. W. Feldman. 1981. Cultural transmission and evolution: a quantitative approach. Princeton University Press, Princeton NJ.

Chaffee, D. W., H. Griffin, and R. T. Gilman. 2013. Sexual imprinting: what strategies should we expect to see in nature? Evolution 67:3588–3599.

Charlesworth, B. 1978. Some models of the evolution of altruistic behaviour between siblings. Journal of Theoretical Biology 72:297–319.

Cornwallis, C. K., S. a. West, K. E. Davis, and A. S. Griffin. 2010. Promiscuity and the evolutionary transition to complex societies. Nature 466:969–972.

Danchin, É., A. Charmantier, F. A. Champagne, A. Mesoudi, B. Pujol, and S. Blanchet. 2011. Beyond dna: integrating inclusive inheritance into an extended theory of evolution. Nature Reviews Genetics 12:475.

Day, T., and R. Bonduriansky. 2011. A unified approach to the evolutionary consequences of genetic and nongenetic inheritance. American Naturalist 178:E18–E36.

Débarre, F., C. Hauert, and M. Doebeli. 2014. Social evolution in structured populations. Nature Communications 5:3409.

Doebeli, M., and N. Knowlton. 1998. The evolution of interspecific mutualisms. Proceedings of the National Academy of Sciences of the USA 95:8676–8680.

Dridi, S., and E. Akcay. 2018. Learning to cooperate: The evolution of social rewards in repeated interactions. American Naturalist 191:58–73.

Dridi, S., and L. Lehmann. 2015. A model for the evolution of reinforcement learning in fluctuating games. Animal Behaviour 104:87–114.

Enquist, M., and S. Ghirlanda. 2005. Neural networks and animal behavior. Princeton University Press, Princeton NJ.

Estrela, S., E. Libby, J. Van Cleve, F. Débarre, M. Deforet, W. R. Harcombe, J. Peña, S. P. Brown, and M. E. Hochberg. 2018. Environmentally mediated social dilemmas. Trends in Ecology & Evolution 34:6–18.

Fawcett, T. W., S. Hamblin, and L.-A. Giraldeau. 2012. Exposing the behavioral gambit: the evolution of learning and decision rules. Behavioral Ecology 24:2–11.

Frank, S. A. 2007. All of life is social. Current Biology 17:R648–R650.

Gardner, A., S. A. West, and N. Barton. 2007. The relation between multilocus population genetics and social evolution theory. American Naturalist 169:207–226.

Godfray, H. C. J. 1995. Signaling of need between parents and young: parent-offspring conflict and sibling rivalry. American Naturalist 146:1–24.

Godfray, H. C. J., G. Parker, and D. Haig. 1991. Clutch size, fecundity and parent-offspring conflict. Philosophical Transactions of the Royal Society B 332:67–79.

Grafen, A. 1984. Natural selection, kin selection and group selection. Pages xx–xx *in* J. R. Krebs and N. B. Davies, eds., Behavioural ecology: an evolutionary approach, 2nd ed.

Hamilton, W. D. 1963. The evolution of altruistic behavior. American Naturalist 97:354–356.

———. 1964. The genetical evolution of social behaviour. Journal of Theoretical Biology 7:1–16.

———. 1972. Altruism and related phenomena, mainly in social insects. Annual Review of Ecology and Systematics 3:193–232.

Hamilton, W. D., and R. M. May. 1977. Dispersal in stable habitats. Nature 269:578–581.

Hauert, C., M. Holmes, and M. Doebeli. 2006. Evolutionary games and population dynamics: Maintenance of cooperation in public goods games. Proceedings of the Royal Society B 273:2565–2570.

Helanterä, H., and T. Uller. 2010. The price equation and extended inheritance. Philosophy and Theory in Biology 2:1–17.

Hilbe, C., Š. Šimsa, K. Chatterjee, and M. A. Nowak. 2018. Evolution of cooperation in stochastic games. Nature 559:246.

Hochberg, M. E., D. J. Rankin, and M. Taborsky. 2008. The coevolution of cooperation and dispersal in social groups and its implications for the emergence of multicellularity. BMC Evolutionary Biology 8:238.

Huang, W., B. Haubold, C. Hauert, and A. Traulsen. 2012. Emergence of stable polymorphisms driven by evolutionary games between mutants. Nature Communications 3:919.

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

Ilany, A., and E. Akçay. 2016. Social inheritance can explain the structure of animal societies. Nature Communications 7:12084.

Ito, K., J. M. McNamara, A. Yamauchi, and A. D. Higginson. 2017. The evolution of cooperation by negotiation in a noisy world. Journal of Evolutionary Biology 30:603–615.

Johnstone, R. A. 1999. Signaling of need, sibling competition, and the cost of honesty. Proceedings of the National Academy of Sciences of the USA 96:12644.

Johnstone, R. A., and C. A. Hinde. 2006. Negotiation over offspring care–how should parents respond to each other's efforts? Behavioral Ecology 17:818.

Kazancıoğlu, E., H. Klug, and S. H. Alonzo. 2012. The evolution of social interactions changes predictions about interacting phenotypes. Evolution 66:2056–2064.

Kerr, B., C. Neuhauser, B. J. Bohannan, and A. M. Dean. 2006. Local migration promotes competitive restraint in a host–pathogen 'tragedy of the commons'. Nature 442:75.

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

Klibaite, U., G. J. Berman, J. Cande, D. L. Stern, and J. W. Shaevitz. 2017. An unsupervised method for quantifying the behavior of paired animals. Physical Biology 14:015006.

Koella, J. C. 2000. The spatial spread of altruism versus the evolutionary response of egoists. Proceedings of the Royal Society B 267:1979–1985.

Krakauer, A. H. 2005. Kin selection and cooperative courtship in wild turkeys. Nature 434:69–72.

Laland, K. N., J. Odling-Smee, W. Hoppitt, and T. Uller. 2013. More on how and why: cause and effect in biology revisited. Biology & Philosophy 28:719–745.

Laland, K. N., K. Sterelny, J. Odling-Smee, W. Hoppitt, and T. Uller. 2011. Cause and effect in biology revisited: is mayr's proximate-ultimate dichotomy still useful? Science 334:1512–1516.

Le Galliard, J.-F., R. Ferriere, and U. Dieckmann. 2005. Adaptive evolution of social traits: origin, trajectories, and correlations of altruism and mobility. American Naturalist 165:206–224.

Lehmann, L. 2007. The evolution of trans-generational altruism: kin selection meets niche construction. Journal of Evolutionary Biology 20:181–189.

———. 2008. The adaptive dynamics of niche constructing traits in spatially subdivided populations: evolving posthumous extended phenotypes. Evolution 62:549–66.

Lehmann, L., and F. Balloux. 2007. Natural selection on fecundity variance in subdivided populations: kin selection meets bet hedging. Genetics 176:361.

Lehmann, L., and M. W. Feldman. 2008. The co-evolution of culturally inherited altruistic helping and cultural transmission under random group formation. Theoretical Population Biology 73:506–516.

Lehmann, L., and L. Keller. 2006a. Synergy, partner choice and frequency dependence: their integration into inclusive fitness theory and their interpretation in terms of direct and indirect fitness effects. Journal of Evolutionary Biology 19:1426–1436.

39

———. 2006*b*. The evolution of cooperation and altruism-a general framework and a classification of models. Journal of Evolutionary Biology 19:1365–1376.

Lehmann, L., N. Perrin, and F. Rousset. 2006. Population demography and the evolution of helping behaviors. Evolution 60:1137–1151.

Lehmann, L., and F. Rousset. 2010. How life history and demography promote or inhibit the evolution of helping behaviours. Philosophical Transactions of the Royal Society B 365:2599–617.

Lewin-Epstein, O., R. Aharonov, and L. Hadany. 2017. Microbes can help explain the evolution of host altruism. Nature Communications 8:14040.

Lion, S., and M. Boots. 2010. Are parasites "prudent"in space? Ecology Letters 13:1245–1255.

Lion, S., and S. Gandon. 2009. Habitat saturation and the spatial evolutionary ecology of altruism. Journal of Evolutionary Biology 22:1487–1502.

Lukas, D., and T. Clutton-Brock. 2012. Cooperative breeding and monogamy in mammalian societies. Proceedings of the Royal Society B 279:2151–2156.

McFarland, D. 1991. Defining motivation and cognition in animals. International Studies in the Philosophy of Science 5:153–170.

McFarland, D. J., and R. M. Sibly. 1975. The behavioural final common path. Philosophical Transactions of the Royal Society B 270:265–93.

McGlothlin, J. W., A. J. Moore, J. B. Wolf, and E. D. Brodie. 2010. Interacting Phenotypes and the Evolutionary Process. III. Social Evolution. Evolution 64:1–17.

McNally, L., S. P. Brown, and A. L. Jackson. 2012. Cooperation and the evolution of intelligence. Proceedings of the Royal Society B 279:3027–3034.

McNamara, J. M. 2013. Towards a richer evolutionary game theory. Journal of the Royal Society Interface 10:20130544.

McNamara, J. M., Z. Barta, L. Fromhage, and A. I. Houston. 2008. The coevolution of choosiness and cooperation. Nature 451:189–192.

McNamara, J. M., C. Gasson, and A. I. Houston. 1999. Incorporating rules for responding into evolutionary games. Nature 401:368–371.

McNamara, J. M., and A. I. Houston. 2009. Integrating function and mechanism. Trends in Ecology and Evolution 24:670–675.

McNamara, J. M., E. M. Wilson, and A. I. Houston. 2006. Is it better to give information, receive it, or be ignorant in a two-player game? Behavioral Ecology 17:441–451.

Montiglio, P.-O., C. Ferrari, and D. Réale. 2013. Social niche specialization under constraints: personality, social interactions and environmental heterogeneity. Philosophical Transactions of the Royal Society B 368:20120343.

Moore, A. J., E. D. Brodie III, and J. B. Wolf. 1997. Interacting phenotypes and the evolutionary process: I. direct and indirect genetic effects of social interactions. Evolution 51:1352–1362.

Mullon, C., L. Keller, and L. Lehmann. 2018. Social polymorphism is favoured by the co-evolution of dispersal with social behaviour. Nature Ecology & Evolution 2:132.

Myerson, R. 1979. Incentive compatibility and the bargaining problem. Econometrica 47:61–74.

Myerson, R., and M. Satterthwaite. 1983. Efficient mechanisms for bilateral trading. Journal of Economic Theory 29:265–281.

Noe, R. 1990. A Veto game played by baboons: a challenge to the use of the Prisoner's Dilemma as a paradigm for reciprocity and cooperation. Animal Behaviour 39:78–90.

———. 2006. Cooperation experiments: coordination through communication versus acting apart together. Animal Behaviour 71:1–18.

Nöldeke, G., and L. Samuelson. 1999. How costly is the honest signaling of need? Journal of Theoretical Biology 197:527–539.

North, D. C. 1991. Institutions. Journal of Economic Perspectives 5:97–112.

Nowak, M. A. 2006. Five rules for the evolution of cooperation. Science 314:1560–3.

Nowak, M. A., C. E. Tarnita, and T. Antal. 2010. Evolutionary dynamics in structured populations. Philosophical Transactions of the Royal Society B 365:19–30.

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

Otto, S. P., F. B. Christiansen, and M. W. Feldman. 1995. Genetic and cultural inheritance of continuous traits. Morrison Institute for Population and Resource Studies Paper Number 64.

Parvinen, K. 2013. Joint evolution of altruistic cooperation and dispersal in a metapopulation of small local populations. Theoretical Population Biology 85:12–19.

Peck, J. R., and M. W. Feldman. 1986. The evolution of helping behavior in large, randomly mixed populations. American Naturalist 127:209–221.

———. 1988. Kin Selection and the Evolution of Monogamy. Science 240:1672–1674.

Pen, I., and P. D. Taylor. 2005. Modelling information exchange in worker–queen conflict over sex allocation. Proceedings of the Royal Society B 272:2403–2408.

Perrin, N., and L. Lehmann. 2001. Is sociality driven by the costs of dispersal or the benefits of philopatry? a role for kin-discrimination mechanisms. American Naturalist 158:471–483.

Pinter-Wollman, N., E. A. Hobson, J. E. Smith, A. J. Edelman, D. Shizuka, S. de Silva, J. S. Waters, S. D. Prager, T. Sasaki, G. Wittemyer, et al. 2013. The dynamics of animal social networks: analytical, conceptual, and theoretical advances. Behavioral Ecology 25:242–255.

Platt, T. G., and J. D. Bever. 2009. Kin competition and the evolution of cooperation. Trends in Ecology & Evolution 24:370–377.

Powers, S. T., A. S. Penn, and R. A. Watson. 2011. The concurrent evolution of cooperation and the population structures that support it. Evolution 65:1527–1543.

Purcell, J., A. Brelsford, and L. Avilés. 2012. Co-evolution between sociality and dispersal: the role of synergistic cooperative benefits. Journal of Theoretical Biology 312:44–54.

Rand, D. G., D. Tomlin, A. Bear, E. A. Ludvig, and J. D. Cohen. 2017. Cyclical population dynamics of automatic versus controlled processing: An evolutionary pendulum. Psychological Review 124:626.

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

Rodrigues, A. M., and H. Kokko. 2016. Models of social evolution: can we do better to pre-

dict 'who helps whom to achieve what'? Philosophical Transactions of the Royal Society B 371:20150088.

Rodríguez-Gironés, M., P. Cotton, and A. Kacelnik. 1996. The evolution of begging: signaling and sibling competition. Proceedings of the National Academy of Sciences of the USA 93:14637.

Rodríguez-Gironés, M., M. Enquist, and P. Cotton. 1998. Instability of signaling resolution models of parent–offspring conflict. Proceedings of the National Academy of Sciences of the USA 95:4453–4457.

Roughgarden, J. 2009. The Genial Gene. University of California Press, Berkeley.

Roughgarden, J., M. Oishi, and E. Akçay. 2006. Reproductive social behavior: cooperative games to replace sexual selection. Science 311:965–969.

Rousset, F. 2004. Genetic Structure and Selection in Subdivided Populations. Princeton University Press, Princeton NJ.

Rousset, F., and S. Billiard. 2000. A theoretical basis for measures of kin selection in subdivided populations: Finite populations and localized dispersal. Journal of Evolutionary Biology 13:814–825.

Roze, D., and F. Rousset. 2008. Multilocus models in the infinite island model of population structure. Theoretical Population Biology 73:529–542.

Saltz, J. B., and B. R. Foley. 2011. Natural genetic variation in social niche construction: social effects of aggression drive disruptive sexual selection in drosophila melanogaster. American Naturalist 177:645–654.

Saltz, J. B., A. P. Geiger, R. Anderson, B. Johnson, and R. Marren. 2016. What, if anything, is a social niche? Evolutionary Ecology 30:349–364.

Sanchez, A., and J. Gore. 2013. Feedback between population and evolutionary dynamics determines the fate of social microbial populations. PLoS biology 11:e1001547.

Servedio, M. R., and M. Kirkpatrick. 1996. The evolution of mate choice copying by indirect selection. American Naturalist 148:848–867.

Sethi, R., and E. Somanathan. 1996. The evolution of social norms in common property resource use. American Economic Review 86:766–788.

Simms, E. L., D. L. Taylor, J. Povich, R. P. Shefferson, J. L. Sachs, M. Urbina, and Y. Tausczik. 2006. An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. Proceedings of the Royal Society B 273:77–81.

Smaldino, P. E., J. C. Schank, and R. McElreath. 2013. Increased costs of cooperation help cooperators in the long run. American Naturalist 181:451–463.

Stamps, J. a. 1991. Why Evolutionary Issues are Reviving Interest in Proximate Behavioral Mechanisms. Integrative and Comparative Biology 31:338–348.

Stewart, A. J., and J. B. Plotkin. 2014. The collapse of cooperation in evolving games. Proceedings of the National Academy of Sciences of the USA 111:17558–17563.

———. 2015. The evolvability of cooperation under local and non-local mutations. Games 6:231–250.

Tarnita, C. E., H. Ohtsuki, T. Antal, F. Fu, and M. A. Nowak. 2009. Strategy selection in structured populations. Journal of Theoretical Biology 259:570–581.

Tavoni, A., M. Schlüter, and S. Levin. 2012. The survival of the conformist: social pressure and renewable resource management. Journal of Theoretical Biology 299:152–161.

Taylor, P. D. 1992. Inclusive fitness in a homogeneous environment. Proceedings of the Royal Society B 249:299–302.

Taylor, P. D., and T. Day. 2004. Stability in negotiation games and the emergence of cooperation. Proceedings of the Royal Society B 271:669–674.

Tilman, A. R., J. Plotkin, and E. Akcay. 2019. Evolutionary games with environmental feedbacks. bioRxiv 493023.

Tilman, A. R., J. R. Watson, and S. Levin. 2017. Maintaining cooperation in social-ecological systems. Theoretical Ecology 10:155–165.

Tinbergen, N. 1963. On aims and methods of ethology. Zeitschrift für Tierpsychologie 20:410–433.

Tramm, N. A., and M. R. Servedio. 2008. Evolution of mate-choice imprinting: competing strategies. Evolution 62:1991–2003.

Trivers, R. L. 1971. The evolution of reciprocal altruism. Quarterly Review of Biology 46:35–57.

Van Baalen, M., and D. A. Rand. 1998. The unit of selection in viscous populations and the evolution of altruism. Journal of Theoretical Biology 193:631–648.

Van Cleve, J. 2015. Social evolution and genetic interactions in the short and long term. Theoretical Population Biology 103:2–26.

———. 2016. Cooperation, conformity, and the coevolutionary problem of trait associations. Journal of Theoretical Biology 396:13–24.

Van Cleve, J., and E. Akçay. 2014. Pathways to social evolution: reciprocity, relatedness, and synergy. Evolution 68:2245–2258.

van den Berg, P., and F. J. Weissing. 2015. The importance of mechanisms for the evolution of cooperation. Proceedings of the Royal Society B 282:20151382.

Van Dyken, J. D., and M. J. Wade. 2012. Origins of altruism diversity II: Runaway coevolution of altruistic strategies via "reciprocal niche construction". Evolution 66:2498–2513.

Vehrencamp, S. L. 1983. Optimal Degree of Skew in Cooperative Societies. American Zoologist 23:327–335.

Vickrey, W. 1961. Counterspeculation, Auctions, and Competitive Sealed Tenders. Journal of Finance 16:8–37.

Wahl, L. M., and M. A. Nowak. 1999. The continuous prisoner's dilemma: I. linear reactive strategies. Journal of Theoretical Biology 200:307–321.

Weitz, J. S., C. Eksin, K. Paarporn, S. P. Brown, and W. C. Ratcliff. 2016. An oscillating tragedy of the commons in replicator dynamics with game-environment feedback. Proceedings of the National Academy of Sciences of the USA 113:E7518–E7525.

West, S. A., A. Griffin, and A. Gardner. 2007. Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. Journal of Evolutionary Biology 20:415–432.

West-Eberhard, M. J. 1989. Phenotypic plasticity and the origins of diversity. Annual review of Ecology and Systematics 20:249–278.

Wey, T., D. T. Blumstein, W. Shen, and F. Jordán. 2008. Social network analysis of animal behaviour: a promising tool for the study of sociality. Animal Behaviour 75:333–344.

47

Whiten, A. 2019. Cultural evolution in animals. Annual Review of Ecology, Evolution, and Systematics 50:1.

Wilson, D. S. 1987. Altruism in mendelian populations derived from sibling groups: the haystack model revisited. Evolution 41:1059–1070.

Wilson, D. S., G. Pollock, and L. A. Dugatkin. 1992. Can altruism evolve in purely viscous populations? Evolutionary ecology 6:331–341.

Wilson, E. O. 1975. Sociobiology: the new synthesis. Harvard University Press.

Worden, L., and S. A. Levin. 2007. Evolutionary escape from the prisoner's dilemma. Journal of Theoretical Biology 245:411–422.

Yeh, D. J., J. W. Boughman, G.-P. Sætre, and M. R. Servedio. 2018. The evolution of sexual imprinting through reinforcement. Evolution 72:1336–1349.

Figure legend

Figure 1: An overview of the components that make up the organization of social interactions. (a) "The game" refers to the description of available strategies (or phenotypes), the sequence of "moves," and each outcome's material consequences for interacting individuals (in terms of resources, reproduction, or survival). (b) "Proximate mechanisms" refer to the mechanistic basis of how genetic variation (represented by the "genotype" arrows going into the brain) is translated into phenotypic expression of behavior, e.g., how individuals make decisions and respond to each other. (c) "Interaction network" captures the patterns of who interacts with whom, in what size groups, and so on. (d) "Demography" captures the demographic and life history structure of the population (e.g., whether there are different classes of individuals) and when important life-history events take place. (e) Finally, "mode of inheritance" refers to how traits get transmitted, which may involve both genetic and non-genetic mechanisms, such as cultural or ecological inheritance. All of these components not only affect selection on social behaviors, but themselves evolve under social selection.
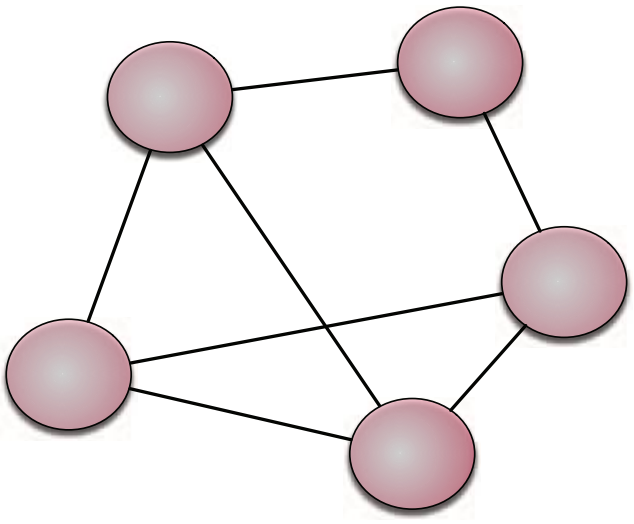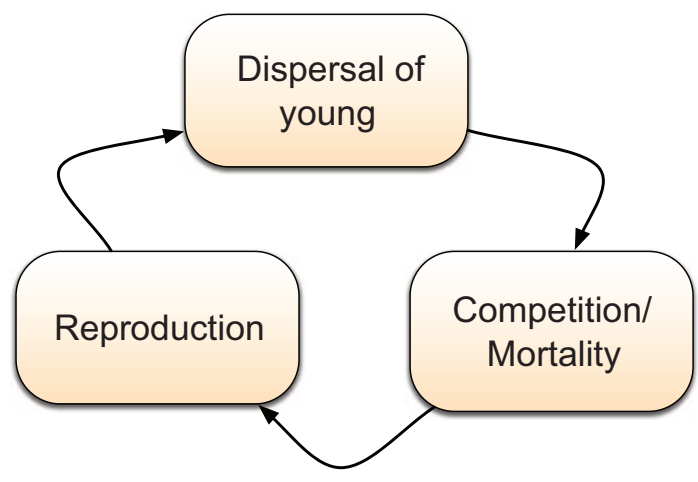
49

Figure 1

(a) the game

(b) proximate mechanisms
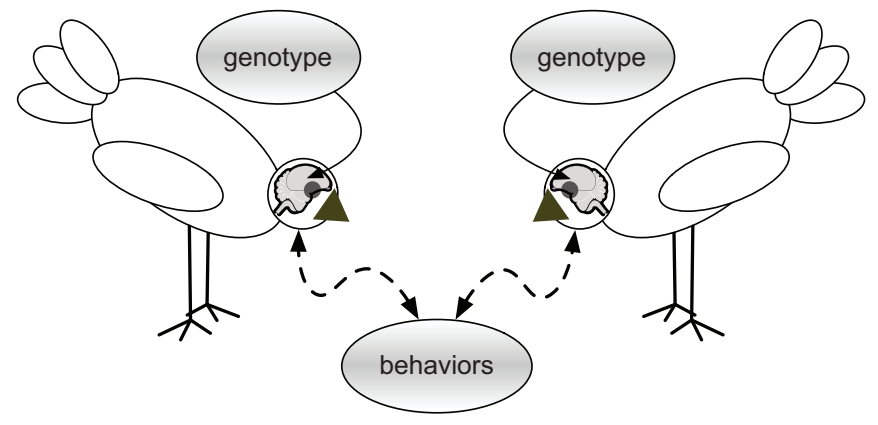
(c) interaction network

(d) demography

(e) mode of inheritance