

# Modeling Social and Evolutionary Games

Angela Potochnik\*

## Abstract

When game theory was introduced to biology, the components of classic game theory models were replaced with elements more befitting evolutionary phenomena. The actions of intelligent agents are replaced by phenotypic traits; utility is replaced by fitness; rational deliberation is replaced by natural selection. In this paper, I argue that this classic conception of comprehensive reapplication is misleading, for it overemphasizes the discontinuity between human behavior and evolved traits. Explicitly considering the representational roles of evolutionary game theory brings to attention areas of overlap that are often neglected, and so a range of evolutionary possibilities that are often overlooked. The clarifications this analysis provides are well illustrated by—and particularly valuable for—game theoretic treatments of the evolution of social behavior.

**keywords:** evolutionary game theory; evolutionary ecology; behavioral ecology; cooperation; altruism

## 1 Game Theory in Biology and Economics

In game theory, interactions are thought of as games, and strategies to playing a game are judged for how they fare in light of other players' strategies. Accordingly, game theory

---

\*Department of Philosophy, University of Cincinnati, 206 McMicken Hall, Cincinnati, OH, 45221-0374. Email: angela.potochnik@uc.edu; Telephone: 513-556-6324

models are applicable whenever elements interact in a way that enables the outcomes of those interactions to affect the properties of the elements. This has yielded, and continues to yield, an astounding range of applications in the social and life sciences.

Game theory was first explicitly introduced by von Neumann & Morgenstern (1944) as a tool in economics. Its classic use was to model the actions of utility-maximizing agents. So, for classic game theory, utility-maximizers are the interacting elements, the elements in a game with one another. The agents' pursuit of greater utility leads them to select a strategy according to its expected utility in light of the circumstances—that is, in light of their expectations of the other agents. This is the means by which outcomes, or expected outcomes, affect the properties of the elements, viz., their strategies.

Maynard Smith & Price (1973) reapplied game theory to account for strategies that result not from rational choice but from evolution by natural selection. In this application, the interacting elements are not rational agents adopting strategies, but organisms with heritable traits. Strategies are thus not chosen for their expected value, but instead evolve in a changing population as a result of their relative fitness (Riechert & Hammerstein, 1983). Accordingly, the need for utility-maximizing rational agents is removed. The game is instead taken to represent the ways in which natural selection shapes the properties of, as Mailath (1992) puts it, “myopic and unsophisticated players.” This reapplication yielded what is known as evolutionary game theory.

The use of evolutionary game theory has proliferated in the decades since its introduction to biology. Any trait that has significance for an organism's survival or reproductive success can be viewed as a strategy, and whenever the success of that trait depends on the traits of other organisms, a game is in play. Whether it pays for a vampire bat to share its hunting spoils with its neighbors depends on whether they will share in return (Wilkinson, 1984). When a male dung beetle should abandon a cowpat for a fresher breeding ground is decided by how long the other males are sticking around and how many new females are still arriving

to the cowpat (Parker, 1970). Game theory has been applied to many other social behaviors, such as issuing warning calls when predators approach and mutualism among members of different species. It has also been used to model relative resistance to parasites, the timing of reproduction in an organism's lifespan, and a long list of other behavioral, physical, and life-history traits. In the process, a vast array of game forms and solution concepts have been developed (McGill & Brown, 2007).

Interestingly, the success of evolutionary game theory in biology has spurred its reapplication in the social sciences, including in areas where classic game theory was already in use. The elements to which evolutionary game theory applies are strategies. This builds an inherent flexibility into the evolutionary process: payoffs might represent the *biological* fitness of organisms or the *cultural* fitness of strategies. Game dynamics that represent biological fitness are often referred to as replicator dynamics (Taylor & Jonker, 1978; Hofbauer & Sigmund, 1998). If strategies are instead influenced by their cultural fitness, this could result in different game dynamics, as determined by different models of learning or imitation (Friedman, 1991). In any case, for evolutionary game theory, the evolutionary process resulting from differences in biological or cultural fitness replace the assumption of rational agents that maximize utility. This is useful even in the social sciences, for there are many circumstances when the assumption of rational agency fails, which could undermine the accuracy of classic game theory models.

This brief survey of game theory's redevelopment for application to evolutionary biology, and then the reimport of evolutionary game theory back into the social sciences, sets the stage for my focus in this paper. The development of evolutionary game theory altered the basic assumptions of game theory in a way that significantly expanded its range of applicability. This is illustrated by evolutionary game theory's success in biology and its successful reapplication to the social sciences. However, at least in biology, the embrace of evolutionary game theory has also often led to the neglect—and even the implicit rejection—

of other, non-evolutionary influences on traits. This is one symptom of a broader difficulty. Evolutionary game theory is generally characterized as the reapplication of the modeling approach to a distinct set of phenomena: selection replaces deliberation, organisms with different phenotypes replace rational agents, and fitness replaces utility. Yet thinking of this as a set of wholesale replacements overemphasizes the discontinuity between the application of game theory to deliberative action and to evolved traits. Indeed, evolutionary game theory's applicability to the social sciences suggests there may be greater continuity between the dynamics of human action and of evolved traits, especially behavior. When that continuity is neglected, the full range of evolutionary possibilities can go unnoticed.

These ideas can be more fully expressed by explicitly considering the representational roles of evolutionary game theory models. I take that approach in §2. There I develop and defend the view that the discontinuity between game theory's application to the social sciences and to evolutionary biology is often overestimated, and I argue that this leads to the neglect of evolutionary possibilities. A prime area of investigation that exemplifies how this is problematic is the evolution of cooperation, so in §3, I discuss the significance of these ideas for game theoretic treatments of the evolution of social behavior.

## 2 How Games Represent Evolution

One way to characterize the reapplication of game theory to evolutionary biology is in terms of what the models are used to represent. As described in the previous section, game theoretic models represent interactions among elements that affect the elements' properties. In classic game theory, the represented elements are individual agents that, by assumption, make decisions based on what will maximize their personal utility. The payoffs are intended to represent the relative utilities of the strategies available to an individual, that is, the agent's potential courses of action. The interactions among agents affect their strategies

game components	represent: rational choice	represent: evolution
players	rational agents	organisms
strategies	individual actions	phenotypes
payoffs	personal utility	lifetime fitness
source of change	deliberation	natural selection
subject to change	agents	population(s)

Table 1: The traditional representational roles of game theory when applied to human decisions versus evolved traits

insofar as the strategies of other agents determine which action will result in maximal utility, and accordingly, which action an agent will choose. The actions of individual agents thus change according to changing utility calculations. For evolutionary game theory, in contrast, the represented elements are organisms, with no assumption of rationality. Interactions among organisms determine the success of strategies, that is, what phenotypes—physical, behavioral, and life history traits—prevail in a population. The interactions exert their influence by affecting the relative fitness of the phenotypes. It is those fitness values that reflect the influence of natural selection. Because natural selection is the source of change, one or more populations are the entities subject to change. Populations, not individuals, have the capacity to evolve. Accordingly, changes in strategy are taken to result from the changing composition of the population instead of the same individuals performing different actions. This conception of the contrasting representational roles of classical and evolutionary game theory is summarized in Table 1.

As indicated in the previous section, the initial reapplication of game theory to such a different set of target systems was a tremendous breakthrough. Evolutionary game theory bypasses the stringent rationality requirements of classic game theory and allows for the application of game theory to a broad range of organismal traits, as well as human behaviors that violate the utility-maximizing assumption. In the first chapter of his classic text on evolutionary game theory, Maynard Smith emphasizes this feature: “it can be a simple

consequence of the laws of population genetics that, at equilibrium, certain quantities are maximised. [...] Nothing is implied about intention...” (Maynard Smith, 1982, p. 5).

Yet the standard conception of evolutionary game theory as the wholesale replacement of one set of representational roles with another has a drawback. It can eclipse the representation of the elements and dynamics that were the focus of classic game theory, and that remain the focus of evolutionary game theory applied to the social sciences. These include (a) strategies that are under individual control and (b) individual game payoffs that do not predictably sum to lifetime fitness values. This may lead some evolutionary biologists to underestimate the significance of these features in the biological world. I address each of these two possibilities in the following subsections.

## 2.1 Questioning Strategies as Evolved Phenotypes

When game theory was redeployed to model evolutionary phenomena, it was natural that the game strategies were taken to represent phenotypes, viz., physical, behavioral, and life-history traits that can be the target of selection. Natural selection, after all, is the source of population-level change at the root of evolutionary game theory. When one strategy has a higher payoff than all others, that phenotype is favored by selection, and is the evolutionary stable strategy (ESS). When no single phenotype is an ESS, Maynard Smith & Price (1973) identify two possibilities, namely,

that an evolutionary stable population is either genetically polymorphic, the strategies of individuals being distributed as in [the equation defining the mixed ESS], or that it consists of individuals whose behaviour differs from contest to contest as in [the equation defining the mixed ESS] (17).

Maynard Smith and Price see the possibilities as (a) a population that has a mixture of genotypes, resulting in a stable distribution of different phenotypes, or (b) a monomorphic

population with a mixed strategy phenotype, viz., a single strategy that yields different results in the face of different game conditions. The focus thus remains on the distribution of phenotypes that (by assumption) results from selection, however that distribution is realized by individual organisms. This focus on evolved phenotypes instead of individual behaviors of individual organisms keeps natural selection center stage.

Yet conceiving of the strategy as an evolved phenotype may eclipse other influences on traits. For many of the phenotypes modeled using evolutionary game theory, there is plausibly a broad range of significant influences besides natural selection. Setting aside the possibility of non-selective evolutionary influences, such as constraints and genetic drift, there is also the possibility that physical traits, behaviors, and aspects of life history have *non-evolutionary* influences. It is increasingly appreciated that development plays a nontrivial role for many traits, and that perhaps the evolution of those traits cannot be understood without reference to the developmental system. Additionally, behavioral traits are particularly good candidates for extensive lability within an individual's lifespan—change in response to environmental influences, perhaps even according to a process of learning from others' behaviors. So it may be that some game strategies change during an organism's lifespan, potentially in complex ways that are not determined by the influence of natural selection. This introduces the possibility that game strategies may not themselves be evolved phenotypes.

Some biologists have recently explored the possibility of within-lifespan influences on traits modeled with evolutionary game theory, especially behavioral traits. Houston & McNamara (1999) advocate greater attention to within-lifespan change than is standard in evolutionary game theory. McNamara et al. (1999) and Taylor & Day (2004) develop evolutionary game theory models that incorporate negotiation, and Taylor et al. (2006) develops a model incorporating phenotypic plasticity, or within-lifespan change. The possibility of within-lifespan change is also at the center of Roughgarden's current research

program (Roughgarden et al., 2006; Roughgarden, 2009). Roughgarden puts the general idea for behavioral traits as follows: “a gene might conceivably code for every type, or even every instance, of behavior. However, an animal’s behavior obviously develops during its life reflecting experience in local situations” (2009, p. 109). This is especially plausible for social behaviors, which by definition involve direct interaction with others and which are particularly amenable to game theoretic treatments.<sup>1</sup>

Roughgarden (2009) illustrates this idea with a simple example of how two birds sharing a nest may allocate their time. By assumption, each bird divides its time between tending the nest and foraging for worms. Which activity is in a bird’s best evolutionary interest depends upon what the other bird is doing, since an unguarded nest is at risk of predation, but unfed offspring risk starvation. Other things equal, it is in a bird’s best interest to stick close to the nest; it is assumed that guarding the nest is safer for the bird than foraging. These relative values are reflected in the payoff matrix in Table 2. For present purposes, the important feature of this example is that a bird’s behavior—how it divides its time between guarding and foraging—is plausibly shaped by its local environment, and especially by the behavior of the bird with which it’s paired. That is, the game dynamics may not be governed by natural selection alone, but a combination of evolved tendencies and direct environmental influences.

	guard	forage
guard	1, 1	4, 2
forage	2, 4	0, 0

Table 2: Payoff matrix for two birds sharing the responsibility of rearing young

---

<sup>1</sup>In what follows, I primarily focus on behavioral traits. This is partly because existing game theory models incorporating within-lifespan influences largely regard behavioral traits and partly because the possibility of within-lifespan change is especially apparent for behavior. Nonetheless, the possibility of significant developmental and environmental influences on physical and life-history traits, as well as behaviors, assures that the point is more generally applicable.

A word of clarification is in order. As I have argued elsewhere, game theory’s focus on natural selection does not require a commitment to selection as the only important influence on the modeled traits (Potochnik, 2009). Game theoretic models may by design only partially represent the target system, in order to focus on a subset of dynamics that are of particular interest. In the same vein, an evolutionary game theory model’s neglect of developmental factors, individual learning, etc.—just like its neglect of genetic drift—does not require the explicit *rejection* of these influences. In order to focus on natural selection’s role in producing a trait, game theory models introduce the idealization that strategies evolve simply according to selective advantage, thereby ignoring any influence of random drift. Similarly, a game theory model may introduce the idealization that strategies themselves are directly subject to selection. This allows for the investigation of the influence of natural selection while other influences on the strategy—such as development or learning—are set aside. My view diverges from Roughgarden’s on this. Roughgarden claims that the explicit treatment of within-lifespan behavioral influences “is an improvement because it eliminates the genetic determinism required in the standard...formulations” (2009, p. 119). In my view, the standard applications of evolutionary game theory should not be interpreted as committed to genetic determinism. These models do not *rule out* other influences on traits; they simply neglect them in order to achieve focused representation of the role of natural selection.

Thus, my criticism is not that evolutionary game theory neglects some of the influences on game outcomes, for this could be the result of a principled focus on natural selection. Instead, I have two more specific concerns. The more basic concern is that some researchers may conclude from evolutionary game theory’s neglect of non-selective influences that there are no such influences on the phenotypes in question, or at least no important influences. That would be a mistake. A model’s neglect of developmental or learning processes should not be taken as evidence that those processes do not exist, or are not significant. They may simply be omitted from a model in order to achieve focused representation of the role of

natural selection. It would be a mistake to assume that, e.g., the behaviors of nesting birds are genetically hardwired, even if a model represents the behaviors as directly subject to selection.

The second concern is that the focused representation of the influence of natural selection may not by itself provide a sufficient treatment of game dynamics. Standard applications of evolutionary game theory represent only the role of natural selection. If a trait is heavily influenced by other factors, such as within-lifespan development, learning, or environment, then standard evolutionary game theory will fail to capture important features of the scenario. These unrepresented influences may go unnoticed. This is particularly so if they too are frequency dependent, since game theory is already in use in a different representational role. In a review of (Roughgarden, 2009), McNamara voices a version of this concern: “it could be argued that behavioral ecologists have previously overemphasized payoffs and paid insufficient attention to the process by which decisions are made, so I am sympathetic to an approach that emphasizes process” (2009, pp. 410, 411).

Further consideration of the example of nesting birds illustrates this point. Roughgarden (2009) points out that natural selection alone acting on the trait guard-or-forage would lead to one of two endpoints: either the female bird will spend all her time foraging and the male will always guard the nest, or vice versa. Both are Nash equilibria and evolutionary stable states. However, another outcome is possible if the birds’ behaviors are shaped by their interactions with one another. Notice from Table 2 that the payoffs to the two birds at the equilibrium points are not equal; the bird that takes on foraging responsibility suffers greater risk of predation. If the birds can communicate—perhaps even offer “threats, promises, and side-payments” (Roughgarden et al., 2006, p. 965)—they can arrange a compromise whereby they trade off foraging responsibility and its attendant risks. This is not to say that natural selection is *irrelevant* to the birds’ behavior. The suggestion is that the simple trait guard-or-forage is not itself an adaptation. Instead, the evolved trait may be a much broader system

of signaling, perhaps including threatening or negotiation, that helps enforce an efficient and fair division of labor.

Possibilities like these are largely neglected if it is assumed that game dynamics are driven by natural selection alone, as often occurs in evolutionary game theory.<sup>2</sup> Developmental and environmental influences, including the direct or indirect influence of others' behaviors, can also influence frequency-dependent characteristics. The possibility of such non-evolutionary influences should not be eclipsed by natural selection; they too warrant investigation.

To summarize, strategies treated by evolutionary game theory are not always best understood as stable, evolved phenotypes. Instead, those strategies may develop within individuals' lifetimes, in response to the environment, including the social environment. Consequently, game dynamics do not always represent natural selection; they may in part or exclusively reflect developmental or learning processes. These possibilities are neglected by standard formulations of evolutionary game theory. It should not be assumed that the successful application of evolutionary game theory indicates the absence of non-evolutionary influences. The possibility of frequency-dependent dynamics within organisms' lifespans should be entertained—especially, but not only, for behavioral traits—and game theory may have an additional application in the treatment of those within-lifespan dynamics.

## 2.2 Questioning Payoffs as Lifetime Fitnesses

There is another respect in which the full range of possibilities goes unrecognized in the standard formulation of evolutionary game theory. This can be appreciated by considering the representational role of game payoffs. Traditionally, payoffs are taken to represent strategies' contribution to lifetime fitness. For example, in their seminal game theory

---

<sup>2</sup>Standard evolutionary game theory can accommodate some within-lifespan influence on traits by considering mixed strategies and response gradients (James Griesemer, in conversation). This facilitates the accurate representation of selection's influence when there is also within-lifespan change, but it usually does not provide a way to represent the dynamics of that change.

treatment of combat among animals of the same species, Maynard Smith and Price define payoffs as follows:

The pay-offs are taken as measures of the contribution the contest has made to the reproductive success of the individual. They take account of three factors: the advantages of winning as compared with losing, the disadvantage of being seriously injured, and the disadvantage of wasting time and energy in the contest (Maynard Smith & Price, 1973, p. 15).

The payoffs are designed to incorporate all of the contest's influences on fitness: by assumption, these are winning versus losing, risk of injury, and the investment of time and energy. Yet I suspect that the payoffs of many games fail to fully represent the contribution to lifetime fitness. It may be that evolutionary game theory is haunted by its history—its origins in classical games—for payoffs often more closely resemble the immediate outcomes of individual encounters than the outcomes' full range of fitness effects.

Of course, there is a relationship between the immediate outcomes of games and a strategy's fitness: the outcomes contribute to overall lifetime fitness. When games' immediate outcomes exhaust their fitness effects, those outcomes can be used as proxies for lifetime fitness. This is what Maynard Smith and Price expect when they calculate payoffs based on the costs/benefits of losing/winning, the risk of injury, and the investment of time and energy. This is similar to how actual reproductive success is often estimated via the stand-in of a single factor that is a component of fitness. For instance, lifetime fitness can be estimated via the number of offspring reared to weaning in a single season, or via the size of a plant's seed set (Lewontin, 1974). However, such proxies for lifetime fitness are conjectural—it may be demonstrated that they do not accurately reflect the full fitness effects of the trait in question. A well-known example of this is that accounting for the evolution of sex ratios requires considering fitness effects in terms of numbers of grand-offspring, not merely offspring. In the same way, the outcomes of individual games may fail

to accurately reflect the full fitness effects of strategies. Games may affect fitness in ways that extend beyond their immediate outcomes.

For an illustration, let us return to Roughgarden's (2009) simple example of two birds caring for their nest of young. The payoff matrix in Table 2 shows the direct fitness consequences of guarding versus foraging; it takes into account the risk of predation and the risk of offspring starvation or predation. As Roughgarden imagines the scenario, an unguarded nest is lost to predation, yielding the lowest relative fitnesses for both the male and female; both guarding protects the nest but some of the nestlings die of starvation. The highest payoffs for both birds occur when one forages and the other guards, though the forager suffers a greater risk of predation. These are the immediate consequences of the behavioral alternatives in question, but there may be other, less obvious consequences. Roughgarden discusses the possibility that a full-time guarder risks punishment by the other bird, which may opt to suffer the decreased fitness of refusing to forage in order to threaten its partner. If this happens with any regularity, it decreases the lifetime fitness of full-time guarders. This is just one possible extra-game fitness effect. The possibilities are boundless: perhaps full-time guarders are less likely to find mates the next season, or perhaps they have partners that are less monogamous than they otherwise would be; perhaps full-time foragers are rewarded with less risky options in other coordinated activities; etc.

To clarify, the idea is not simply that games may be more complicated than expected, or may have different payoffs than the ones in the model. Instead, the point is that strategies may have regular (but not apparent) downstream consequences that significantly influence their lifetime fitness effects. A model may correctly represent the immediate consequences of a strategy that affect fitness, but incorrectly assume that these immediate consequences accurately reflect *lifetime* fitness effects. This is related to the possibility discussed in Section 2.1 that developmental or learning dynamics may in part determine the outcomes of games. It is primarily the possibility of changing environmental and social conditions due

to the strategy employed that allows lifetime fitness effects to diverge from immediate game consequences. A notable example is the possibility of rewards and punishments that occur in aspects of life other than the interaction represented in the model.

A prime example of evolutionary game theory models employing payoffs that do not represent lifetime fitness values is found in models of reciprocal altruism. The evolutionary significance of reciprocal altruism was first modeled in terms of the prisoner’s dilemma by Trivers (1971); Axelrod & Hamilton (1981) developed models that employed the solution concept of an evolutionary stable strategy. The prisoner’s dilemma is distinguished by the relation among payoffs. The players do better if both cooperate than if both defect, but each player improves its payoff by defecting, regardless of whether the other cooperates or defects; see Table 3. The prisoner’s dilemma has been used to model of a variety of cooperative scenarios, many of which rely on reciprocation for the continuance of cooperative behavior. These include the symbiosis between cleaner fish and host fish (Trivers, 1971); birds’ warning calls when predators are near (Axelrod & Hamilton, 1981); and blood-sharing among vampire bats (Wilkinson, 1984), to name only a few.

	cooperate	defect
cooperate	$R, R$	$S, T$
defect	$T, S$	$P, P$

Table 3: Payoff matrix for the prisoner’s dilemma, where  $T > R > P > S$

The use of the iterated prisoner’s dilemma model to represent reciprocal altruism is a clear instance of payoffs that reflect immediate game consequences in a way that diverges from lifetime fitness effects. The iterations combine with other assumptions regarding ability to punish (e.g., the famous strategy of Tit-for-Tat), structured populations, etc., to determine the fitness consequences of the game. In cases of reciprocal altruism, these circumstances turn the within-generational prisoner’s dilemma into something like an evolutionary stag

hunt—i.e., for lifetime fitness, it is best to cooperate if the other player(s) cooperate, best to defect if the other(s) defect as well (Skyrms, 2003). Representing the cumulative effects on lifetime fitness removes the apparent advantages of defection. As West et al. (2007) point out, reciprocal altruists are not truly evolutionary altruists.

In cases of reciprocal altruism, immediate game consequences are not equivalent to lifetime fitness, but they do predictably determine fitness when combined with other assumptions. Still, the failure to carefully distinguish immediate consequences from lifetime fitness effects has led to the neglect of two other sets of possibilities. First, as reciprocal altruism builds punishment directly into the game iterations, there is also the possibility of rewards and punishments—and threats of such—that do not occur within the limited game framework. This possibility was illustrated above with the help of Roughgarden’s bird example. Punishment might take the form of refusal to cooperate in the modeled interaction, and reward the form of continued cooperation, but these might instead involve a host of other social consequences, costly or beneficial. Second, there is the additional possibility of a true *evolutionary* prisoner’s dilemma, that is, lifetime fitness effects that are ordered according to the payoff matrix of a prisoner’s dilemma. Most treatments of the prisoner’s dilemma and the evolution of cooperation do not address this possibility. One possible exception is a model developed by Akçay et al. (2009), which suggests that “other-regarding” motivations—the motivation to increase partners’ payoffs as well as one’s own—can evolve in a range of conditions, and can lead to spontaneous cooperation among unrelated individuals when direct reciprocity is unlikely.

To summarize, though the standard formulation of evolutionary game theory indicates that payoffs represent fitness effects, evolutionary game theory models often employ a payoff structure that represents immediate game consequences, and these can fail to accurately reflect the full consequences for lifetime fitness. This can have several repercussions. There may be a failure to distinguish evolutionary influences, viz. fitness effects, from within-

lifespan influences such as non-random interactions. This can lead the type of evolutionary game in play to be misdiagnosed, as seen with reciprocal altruism models. Lastly, other influences on game dynamics might go unrecognized, such as rewards, punishments, and threats.

In this section and the last (§2.1, 2.2) I have outlined additional representational roles that evolutionary game theory may have. These additional roles are reminiscent of classical game theory, insofar as they deal with the actions of individuals and within-lifespan changes. In Table 4 I detail the traditional representational roles of evolutionary game theory, alongside the additional alternatives that I have identified. Notice that strategies may best represent either stable phenotypes or what I have termed “game consequences.” These game consequences may be components of lifetime fitness, or they may not influence biological fitness at all. Accordingly, sources of change may include development or learning, in addition to or instead of natural selection, and individual organisms may change within their lifespans, in addition to or instead of change in the composition of the population over generations. Modeling evolutionary games with an eye to this range of representational possibilities will help avoid ambiguities and neglected possibilities, and it will highlight the full scope of possible influences on organismal traits.

game components	representational possibilities	
players		organisms
strategies	individual actions	stable phenotypes
payoffs	game consequences	lifetime fitness
sources of change	development/learning (frequency dynamics)	natural selection (population dynamics)
subject to change	individuals	population(s)

Table 4: An increased range of possibilities for game theory’s representation of evolved phenotypes

### 3 Modeling the Evolution of Social Behavior

The expanded set of representational possibilities for evolutionary game theory identified in §2 is particularly apt for the evolution of social behavior. It is well appreciated that, as Lewontin put it, “organisms do not experience environments passively; they create and define the environment in which they live” (1978, p. 215). This is especially true of organisms’ *social* environments. Evolutionary game theory is a brilliant reapplication of rational choice theory to the blind action of natural selection, but it should not be overlooked that the terms of the games may evolve in tandem with strategic interactions (Calcott, 2008). This enables immediate game consequences to have different impacts on lifetime fitness, and it allows strategies to be shaped within organisms’ lifespans. Here I discuss some implications of this for modeling the evolution of social behavior.

The rich social lives of many animal species provide the opportunity for a variety of structured interactions, and those interactions may not be discrete. As discussed above, this can allow for interwoven social behaviors that are potentially shaped by individual experience. For social behaviors especially, the dynamics of within-lifespan change, viz. frequency dynamics, may also be amenable to game theoretic treatment. It is thus important to distinguish sources of behavioral change from sources of evolutionary change; the model of other-regard developed by Akçay et al. (2009) exemplifies this approach for the evolution of cooperation. Explicitly distinguishing between behavioral and evolutionary dynamics provides resources for a more nuanced understanding of cooperative or altruistic behavior. Indeed, different definitions of altruism and cooperation are relevant in these two contexts. Here I will focus on cooperation, as this is generally taken to be the more inclusive term.<sup>3</sup>

The common evolutionary definition of cooperation is in terms of fitness effects: a behavior is considered to be cooperative when it benefits—i.e., increases the fitness of—

---

<sup>3</sup>See West et al. (2007) for an account of the different usages of “cooperation,” “altruism,” and related terms, and the confusion surrounding them.

others. A cooperative behavior in this sense may also benefit the actor, or it may be at the cost of the actor's direct fitness. The latter is the common evolutionary definition of altruism. This definition of cooperation describes a type of *evolutionary* dynamics—more precisely, population dynamics driven by selection. Most are careful to distinguish this evolutionary sense of cooperation from the psychological sense, according to which cooperative behaviors are those performed out of regard for the well-being of others (e.g. Sober & Wilson, 1998). The latter definition of cooperation describes a type of *psychological* dynamics: what motivates an individual's single act or behavioral trend. The distinction between psychological and evolutionary cooperation can be used as a template for a further definition of cooperation, describing a type of *behavioral* dynamics. Cooperation in this sense would be defined as actions that immediately benefit others. Consider once more Roughgarden's simple model of nesting birds. Cooperative actions in the behavioral sense include a bird choosing to forage when its partner has chosen to guard the nest, or vice versa; refraining from enacting an established threat, or extending a reward. A sense of behavioral altruism follows: an altruistic action immediately benefits others at some immediate cost to the individual. Many reciprocal altruism models describe altruism in this sense, not evolutionary altruism (West et al., 2007).

This behavioral sense of cooperation is similar to Roughgarden's (2006; 2009) proposal for defining cooperation, according to which a cooperative behavior is “a process of perceiving and playing the game” whereby players act in the interest of team fitness, rather than individual fitness. Yet there are two main differences. First, Roughgarden advocates redefining cooperation in this way, while I think it is important to maintain the existing common definition of evolutionary cooperation, and to define behavioral cooperation in a parallel way. The two senses of cooperation can be distinguished by which type of dynamics are described. The distinction is evident when behavioral and evolutionary dynamics are explicitly treated, as advocated by Roughgarden and in §2 above. There is some precedent

for maintaining three distinct senses of altruism that parallel my suggested distinction among senses of cooperation (Kitcher, 2010). A second difference between Roughgarden's proposed behavioral sense of cooperation and mine is in the details. I worry that the requirement that players perceive and play the game according to team fitness is too strong. This requirement seems to stray into the territory of psychological dynamics, and it accordingly limits behavioral cooperation to psychologically developed organisms that can perceive and consciously act (cf. Milam et al., 2011). Explicitly defining behavioral cooperation in terms of organisms' actions avoids this limitation, and it maintains the tripartite distinction among psychological, behavioral, and evolutionary dynamics.

All three senses of cooperation are potentially of evolutionary relevance. As for the psychological sense of cooperation, I have already mentioned the model by Akçay et al. (2009), showing conditions in which regard for others can evolve, and how this other-regard can enable the evolution of cooperation among unrelated individuals. Other-regard describes a variety of motivation, so it is a form of psychological cooperation proposed to be of evolutionary significance. More generally, McNamara (2009) points out that, although "previous work in [evolutionary biology] has often ignored the mental machinery by which decisions are reached...The fact that decisions are often mediated by high-level mental states such as anger or trust may well affect the behavioral strategies that evolve" (p. 411). McNamara et al. (1999) and Akçay et al. (2009) begin to explore such possibilities.

As for the behavioral sense of cooperation, Roughgarden (2009) posits that behavioral cooperation can influence evolution by removing the evolutionary cost of cooperation for individuals. That is, behavioral dynamics can develop that align an individual's best interest with others' best interests, thereby eliminating the need for evolutionary altruism. Indeed, Worden & Levin (2007) demonstrate the possibility of a population evolving away from the prisoner's dilemma to a payoff structure that removes the dilemma. This can happen through the emergence of systems of rewards, punishments, and threats. Ross (2010) says of

the prisoner’s dilemma that “it is the logic of the prisoners’ situation, not their psychology, that traps them in the inefficient outcome. . . Agents who wish to avoid inefficient outcomes are best advised to prevent certain games from arising.” Ross poses this as a suggestion for rational choice, but there is an analog for evolutionary games. Natural selection can prevent certain evolutionary games from arising, via the evolution of systems of behavioral dynamics that remove the conflict in fitness consequences.<sup>4</sup>

The recognition of the potential evolutionary significance of psychological and behavioral dynamics—and the potential variety of representational roles for evolutionary game theory—has implications for modeling approaches in evolutionary biology. These motivate the explicit consideration of the distinct sets of evolutionary, behavioral, and psychological dynamics. One alternative is the development of models that simultaneously represent the different sets of dynamics; this is the option pursued by Roughgarden et al. (2006), Roughgarden (2009), and Akçay et al. (2009). Another option is the continued use of traditional evolutionary game theory models, but with more explicit consideration of what features the model represents, and what features it neglects. This approach would be in line with how most or all approaches to modeling evolution, including evolutionary game theory, neglect an array of influences in order to represent focal dynamics (Potochnik, 2009, 2010; §2.1 above).

As surveyed at the beginning of this paper, game theory has benefitted tremendously from cross-pollination among its applications in the social sciences and in evolutionary biology. At its most basic, my suggestion in this paper is that many standard formulations of evolutionary game theory in biology overemphasize the discontinuity between these applications. Many evolved traits, and especially social behaviors, may be influenced by rich behavioral and/or psychological dynamics that are distinct from—and important to—the evolutionary dynamics. Careful attention to the possible representational roles of

---

<sup>4</sup>This possibility is investigated by Akçay et al. (unpublished) using the political and economic game theory concept of mechanism design.

evolutionary game theory illuminates the wide variety of evolutionary possibilities, and how they may interface with behavioral and psychological possibilities. For the evolution of social behavior, this brings to light a much broader variety of ways in which cooperation can emerge, facilitated by a range of possible relations among psychological dynamics, behavioral dynamics, and evolutionary dynamics.

## Acknowledgments

This project was influenced by the Function and Evolution Working Group at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation, the U.S. Department of Homeland Security, and the U.S. Department of Agriculture through NSF Award #EF-0832858, with additional support from the University of Tennessee, Knoxville. I received inspiration and helpful input from the members of that working group, especially Erol Akçay, Joel Brown, James Fearon, Lou Gross, Joan Roughgarden, and Lee Worden. I also received helpful feedback from the members of the Philosophical Pizza Munch at the California Academy of the Sciences, including especially Elihu Gerson, James Griesemer, and Roberta Millstein; from Staffan Bergwik, Samir Okasha, and Michael Weisberg; and from three anonymous referees. Some of this work was completed during a research fellowship at the Michelle R. Clayman Institute for Gender Research.

## References

- Akçay, E., Roughgarden, J., Fearon, J., Ferejohn, J., & Weingast, B. (unpublished).  
Biological institutions: The political science of animal cooperation.
- Akçay, E., van Cleve, J., Feldman, M. W., & Roughgarden, J. (2009). A theory for the

- evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences*, 106(45), 19061–19066.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211(4489), 1390–1396.
- Calcott, B. (2008). The *other* cooperation problem: Generating benefit. *Biology and Philosophy*, 23, 179–203.
- Friedman, D. (1991). Evolutionary games in economics. *Econometrica*, 59(3), 637–666.
- Hofbauer, J., & Sigmund, K. (1998). *Evolutionary Games and Replicator Dynamics*. Cambridge: Cambridge University Press.
- Houston, A. I., & McNamara, J. M. (1999). *Models of Adaptive Behaviour*. Cambridge: Cambridge University Press.
- Kitcher, P. (2010). Varieties of altruism. *Economics and Philosophy*, 26(2), 121–148.
- Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. New York: Columbia University Press.
- Lewontin, R. C. (1978). Adaptation. *Scientific American*, 239, 212–230.
- Mailath, G. J. (1992). Introduction: Symposium on evolutionary game theory. *Journal of Economic Theory*, 57, 259–277.
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, 246, 15–18.
- McGill, B. J., & Brown, J. S. (2007). Evolutionary game theory and adaptive dynamics of continuous traits. *Annual Review of Ecology, Evolution and Systematics*, 38, 403–435.

- McNamara, J. M. (2009). Review of *The Genial Gene: Deconstructing Darwinian Selfishness*. *The Quarterly Review of Biology*, 84, 410–411.
- McNamara, J. M., Gasson, C. E., & Houston, A. I. (1999). Incorporating rules for responding into evolutionary games. *Nature*, 401, 368–371.
- Milam, E., Millstein, R., Potochnik, A., & Roughgarden, J. (2011). Sex and sensibility: The role of social selection, a review symposium of roughgarden’s *The Genial Gene*. *Metascience*, 20(2), 253–277.
- Parker, G. (1970). The reproductive behavior and the nature of sexual selection in *Scatophaga stercoraria* L. (*Diptera: Scatophagidae*): Ii. The fertilization rate and the spatial and temporal relationships of each sex around the site of mating and oviposition. *The Journal of Animal Ecology*, 39(1), 205–228.
- Potochnik, A. (2009). Optimality modeling in a suboptimal world. *Biology and Philosophy*, 24(2), 183–197.
- Potochnik, A. (2010). Explanatory independence and epistemic interdependence: A case study of the optimality approach. *The British Journal for the Philosophy of Science*, 61(1), 213–233.
- Riechert, S. E., & Hammerstein, P. (1983). Game theory in the ecological context. *Annual Review of Ecology and Systematics*, 14, 377–409.
- Ross, D. (2010). Game theory. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. forthcoming URL = <http://plato.stanford.edu/archives/fall2010/entries/game-theory/>, fall 2010 ed.
- Roughgarden, J. (2009). *The Genial Gene: Deconstructing Darwinian Selfishness*. Berkeley: University of California Press.

- Roughgarden, J., Oishi, M., & Akçay, E. (2006). Reproductive social behavior: Cooperative games to replace sexual selection. *Science*, *311*, 965–969.
- Skyrms, B. (2003). *The Stag Hunt and the Evolution of Social Structure*. Cambridge: Cambridge University Press.
- Sober, E., & Wilson, D. S. (1998). *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.
- Taylor, P. D., & Day, T. (2004). Stability in negotiation games and the emergence of cooperation. *Proceedings of the Royal Society of London, Series B*, *271*(1540), 669–674.
- Taylor, P. D., Day, T., Nagy, D., Wild, G., André, J.-B., & Gardner, A. (2006). The evolutionary consequences of plasticity in host-pathogen interactions. *Theoretical Population Biology*, *69*, 323–331.
- Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, *40*, 145–156.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press.
- West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology*, *20*, 415–432.
- Wilkinson, G. S. (1984). Reciprocal food sharing in the vampire bat. *Nature*, *308*, 181–184.

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