1 Introduction

A male peacock mantis shrimp resides happily in his burrow. He has recently molted, leaving his shell soft and vulnerable. As he waits, another male wanders onto his territory and approaches. The new male would like a burrow of his own. Both possess raptorial claws powerful enough to smash through the glass of an aquarium. If they fight for the territory, the temporarily squishy burrow owner will be seriously hurt. Neither, though, can directly observe whether the other has molted recently. Both mantis shrimp raise their appendages to display brightly colored "meral spots", intended to signal their aggression and strength. The intruder is impressed and backs off to seek territory elsewhere.

Alex is looking to hire a new data analyst for his company. Annaleigh wants the job and so is keen to impress Alex. She includes every accomplishment she can think of on her resume. Since she went to a prestigious college, she makes sure her educational background is front and center, where Alex will see it first.

A group of vampire bats need to eat nightly to maintain their strength. This is not always easy, however. Sometimes a bat will hunt all night but fail to find a meal. For this reason, most bats in the group have established relationships for reciprocal food sharing. Bats who manage to feed will regurgitate blood for partners who did not.

Mitzi needs a kidney transplant but is too old to go on the normal donor list. Her close friend wants to donate but is incompatible as a donor for Mitzi. They join the National Kidney Registry and become part of the longest-ever kidney donation chain. This involves thirty-four donors who pay forward kidneys to unrelated recipients so that their loved ones receive donations from others in the chain.[1]

We have just seen four examples of strategic scenarios. By strategic, I mean situations where (1) multiple actors are involved and (2) each actor is affected by what the others do. Two of these strategic scenarios I just described – the ones involving mantis shrimp and vampire bats – occurred between non-human animals. The other two – with Alex, Annaleigh, Mitzi, and her friend – occurred in the human social scene. Notice that strategic scenarios are completely ubiquitous in both realms. Whenever predators hunt prey, potential mates choose partners, social animals care for each other, mutualists exchange reciprocal goods or services, parents feed their young, or rivals battle over territory, these animals are engaged in strategic scenarios. Whenever humans do any of these things, or when they apply for jobs, work together to overthrow a dangerous authoritarian leader, bargain over the price of beans, plan a military

[1] This one is a true story. See UW Health (2015).

attack, divide labor in a household, or row a boat, they are likewise in strategic scenarios.

This ubiquity explains the success of the branch of mathematics called *game theory*. Game theory was first developed to explore strategic scenarios among humans specifically.[2] Before long, though, it became clear that this framework could be applied to the biological world as well. There have been some differences to how game theory has typically been used in the human realm versus the biological one. Standard game theoretic analyses start with a *game*. I will define games more precisely in Section 2, but for now it is enough to know that they are simplified representations of strategic scenarios. Game theoretic analyses typically proceed by making certain assumptions about behavior, most commonly that agents act in their own best interests and that they do this rationally.

Strong assumptions of rationality are not usually appropriate in the biological realm. When deciding whether to emit signals to quora of their peers, bacteria do not spend much time weighing the pros and cons of signaling or not signaling. They do not try to guess what other bacteria are thinking about and make a rational decision based on these calculations. But they do engage in strategic behavior, in the sense outlined above, and often very effectively.

There is another framework, closely related to game theory, designed to model just this sort of case. *Evolutionary game theory* also starts with games but focuses less on rationality.[3] Instead, this branch of theorizing attempts to explain strategic behavior in the light of evolution. Typical work of this sort applies what are called *dynamics* to populations playing games. Dynamics are rules for how strategic behavior will change over time. In particular, dynamics often represent evolution by natural selection. These models ask, in a group of actors engaged in strategic interactions, which sorts of behaviors will improve fitness? Which will evolve?

This is an Element about games in the philosophy of biology. The main goal here is to survey the most important literature using game theory and evolutionary game theory to shed light on questions in this field. Philosophy of biology is a subfield in philosophy of science. Philosophers of science do a range of things. Some philosophy of science asks questions like, How do scientists create knowledge? What are theories? And how do we shape an ideal science? Other philosophers of science do work that is continuous with the theoretical

[2] Von Neumann and Morgenstern (1944) originated the theory. Other very influential early contributions were made by John Nash (1950).

[3] This framework was first developed by Maynard-Smith and Price (1973). Before this, though, game theorists were already starting to think about dynamical approaches to games, as in Robinson (1951) and Brown (1951).

sciences, though often with a more philosophical bent. This Element will focus on work in this second vein.

In surveying this literature, then, it will not be appropriate to draw hard disciplinary boundaries. Some work by biologists and economists has greatly contributed to debates in philosophy of biology. Some work by philosophers of biology has greatly contributed to debates in the behavioral sciences. Instead, I will focus on areas of work that have been especially prominent in philosophy of biology and where the greatest contributions have been made by philosophers.

As I've also been hinting at, it will also not be appropriate to draw clear boundaries between biology and the social sciences in surveying this work. This is in part because game theory and evolutionary game theory have been adopted throughout the behavioral sciences – human and biological. Frameworks and results have been passed back and forth between disciplines. Work on the same model may tell us both about the behaviors of job seekers with college degrees and the behaviors of superb birds of paradise seeking mates, meaning that one theorist may contribute to the human and biological sciences at the same time.

There is something special that philosophers of biology have tended to contribute to this literature. Those trained in philosophy of science tend to think about science from a meta-perspective. As a result, many philosophers of science have both used formal tools and critiqued them at the same time. This can be a powerful combination. It is difficult to effectively critique scientific practice without being deeply immersed in that practice, but scientists are not always trained to turn a skeptical eye on the tools they use. This puts philosophers of biology who actually build behavioral models into a special position. As I will argue, this is a position that any modeler should ultimately assume – one of using modeling tools and continually assessing the tools themselves. Throughout this Element we will see how philosophers who adopted this sort of skeptical eye have helped improve the methodological practices of game theory and evolutionary game theory.

So, what are the particular debates and areas of work that will be surveyed here? I will focus on two branches of literature. I begin this Element with a short introduction to game theory and evolutionary game theory. After that, I turn to the first area of work, which uses *signaling games* to explore questions related to communication, meaning, language, and reference. There are three sections in this part of the Element. The first, Section 3, starts with the common interest signaling game described by David Lewis. This section shows how philosophers of science have greatly expanded theoretical knowledge of this model, while simultaneously using it to explore questions related to human and animal signaling. In Section 4, I turn to a variation on this model – the conflict of interest signaling game. This game models scenarios where actors communicate but do not always seek the same ends. As we will see, traditional approaches to conflict of interest signaling, which appeal to signal costs, have been criticized as biologically unrealistic and methodologically unsound. Philosophers of biology have helped to rework this framework, making clearer how costs can and cannot play a role in supporting conflict of interest signaling. The last section on signaling, Section 5, turns to deeper philosophical debates. I look at work using signaling games to consider what sort of information exists in biological signals. Do these signals include semantic content? Of what sort? Can they be deceptive?

The second part of the Element addresses a topic that has been very widely studied in game theory and evolutionary game theory – prosociality. By prosociality, I mean strategic behavior that generally contributes to the successful functioning of social groups. Section 6 looks at the evolution of altruism in the prisoner's dilemma game. This topic has been explored in biology, economics, philosophy, and all the rest of the social sciences, so this section is even more interdisciplinary than the rest of the Element. Because this game *has* been so widely studied, and the literature so often surveyed, I will keep this treatment very brief. Section 7 turns to two models that have gotten quite a lot of attention in philosophy of biology: the stag hunt and the Nash demand game. The stag hunt represents situations where cooperation is risky but mutually beneficial. The Nash demand game can represent bargaining and division of resources more generally. As we will see, these games have been used to elucidate the emergence of human conventions and norms governing things like the social contract, fairness, and unfairness.

By way of concluding the Element, I'll dig a bit deeper into a topic mentioned above: methodology. In particular, a number of philosophers have made significant methodological contributions to evolutionary game theory that stand free of particular problems and applications. The epilogue, Section 8, briefly discusses this literature.

Readers should note that the order and focus of this Element are slightly idiosyncratic. A more traditional overview might start with the earliest work in evolutionary game theory – looking at the hawk-dove game in biology – and proceed to the enormous literature on cooperation and altruism that developed after this. The literature on signaling games would come later, reflecting the fact that it is relatively recent, and would take up a less significant chunk of the Element. My goal is to give a more thorough overview of work that has received less attention rather than to revisit well-worn territory. In addition, as mentioned, I focus in this Element on the areas of the literature that have been most heavily developed in philosophy.

This Element is very short. For this reason, it does not provide a deep understanding of any one topic but overviews many. Interested readers should, of course, use the references in this Element to dig deeper. Furthermore, the brevity of the Element means there is no space to even briefly touch on a number of related literatures. Most notably, I ignore important work in biology and philosophy of biology on other sorts of evolutionary theory/modeling and work in economics on strategic behavior in the human realm.

So, in the interest of keeping it snappy, on to Section 2, and the introduction to game theory and evolutionary game theory.

2 Games and Dynamics

What does a strategic scenario involve? Let's use an example to flesh this out. Every Friday, Alice and Sharla like to have coffee at Peets. Upon arriving slightly late, though, Sharla notices it is closed and Alice is not there. Alice has likely chosen another coffee shop – either the nearby Starbucks or the Dunkin' Donuts. Which should Sharla try?[4] This sort of situation is often referred to as a coordination problem. Two actors would like to coordinate their action, in this case by choosing the same coffee shop. They care quite a lot about going to the same place, and less about which place it is.

How can we formalize this scenario into a game? We do this by specifying four things. First, we need to specify the *players*, or who is involved. In this case, there are two players, Alice and Sharla. If we like, we can call them player 1 and player 2. Second, we need to specify what are called *strategies*. Strategies capture the behavioral choices that players can make. In this case, each player has two choices, to go to Starbucks or to go to Dunkin' Donuts. Let's call these strategies *a* and *b*. Third, we need to specify what each player gets for different combinations of strategies, or what her *payoffs* are.

This is a bit trickier. In reality, the payoff to each player is the joy of meeting a friend should they choose the same strategy and unhappiness over missing each other if they fail. But to formalize this into a game, we need to express these payoffs mathematically. What we do is choose somewhat arbitrary numbers meant to capture the *utility* each player experiences for each possible outcome. Utility is a concept that tracks joy, pleasure, preference, or whatever it is that

[4] Let's suppose neither owns a cell phone.

		Player 2	
		A	B
Player 1	A	1,1	0,0
	B	0,0	1,1

Figure 2.1 A payoff table of a simple coordination game. There are two players, each of whom chooses *a* or *b*. Payoffs are listed with player 1 first.

players act to obtain.[5] Let us say that when Sharla and Alice meet, they get payoffs of 1, and when they do not, they get lower payoffs of 0.

We can now specify a game representing the interaction between Alice and Sharla. Figure 2.1 shows what is called a *payoff table* for this game. Rows represent Alice's choices, and columns represent Sharla's. Each entry shows their payoffs for a combination of choices, with Alice's (player 1's) payoff first. As specified, when they both pick *a* or both pick *b*, they get 1. When they miscoordinate, they get 0. Game theorists call games like this – ones that represent coordination problems – *coordination games*. This game, in particular, we might call a pure coordination game. The only thing that matters to the actors from a payoff perspective is whether they coordinate.

There is one last aspect of a game that often needs to be specified, and that is *information*. Information characterizes what each player knows about the strategic situation – does she know the structure of the game? Does she know anything about the other player? In evolutionary models the information aspect of games is often downplayed, though. This is because information in this sense is most relevant to rational calculations of game theoretic behavior, not to how natural selection will operate.

So now we understand how to build a basic game to represent a strategic scenario. The next question is, how do we analyze this game? How do we use this model to gain knowledge about real-world strategic scenarios? In classic game theory, as mentioned in the introduction, it is standard to assume that each actor attempts to get the most payoff possible given the structure of the game and what the player knows. This assumption allows researchers to derive

5 Utility is a controversial concept. It has been criticized as leading to circular reasoning – players act to get utility because utility is the sort of thing players act to get (Robinson, 1962). In evolutionary models, as we will see, this circularity is less of a worry. A traditional justification of the utility concept in economics stems from Von Neumann and Morgenstern (1944) (this justification was outlined in the second, 1947 edition of their book). They show that agents who satisfy four reasonable axioms have preferences that can be respresented by a utility function. And if these agents act to satisfy these preferences, they will act as if they are maximizing expected utility.

predictions about which strategies a player will choose, or might choose, and also to explain observed behavior in strategic scenarios.

More specifically, these predictions and explanations are derived using different solution concepts, the most important being the *Nash equilibrium* concept. A Nash equilibrium is a set of strategies where no player can change strategies and get a higher payoff. There are two Nash equilibria in the coordination game – both players choose *a* or both choose *b*.[6] Although this is far from the only solution concept in use by game theorists, it is most used by philosophers of biology, who tend to focus on evolutionary models. As we will see, Nash equilibria have predictive power when it comes to evolution, as well as to rational choice–based analyses.[7]

Notice that there is something in the original scenario with Sharla and Alice that is not captured in the game depicted in Figure 2.1. Remember that we supposed Sharla arrived at Peets late, to discover Alice had already made a choice. The payoff table above displays what is called the *normal form* coordination game. Normal form games do not capture the fact that strategic decisions happen over time. Sometimes one cannot appropriately model a strategic scenario without this time progression. In such cases, one can build a game in *extensive form*. Figure 2.2 shows this. Play starts at the top node. First Alice (player 1) chooses between Starbucks and Dunkin' Donuts (*a* or *b*). Then Sharla (player 2)

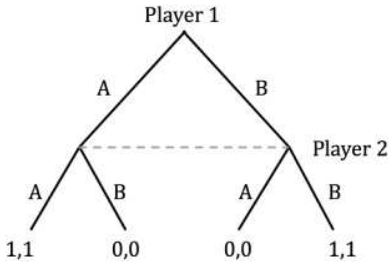

Figure 2.2 The extensive form of a simple coordination game. There are two players, each of whom chooses *a* or *b*. Player one chooses first.

6 This is imprecise. In fact, there are only two *pure strategy* Nash equilibria. These are equilibria in strategies where agents always take the same behavior. At some points in the Element we will look at *mixed strategies*, where agents probabilistically mix behaviors. Many games have Nash equilibria in mixed strategies, and some games only have Nash equilibria in mixed strategies, though we will only discuss these when they are evolutionarily relevant.

7 This predictive power should not be taken too strongly. Experimental evidence finds that in some cases, laboratory subjects play Nash equilibria in the lab, and in other cases not (Smith, 1994). For instance, Güth et al. (1982) give a famous example of the failure of Nash equilibrium predictions to account for bargaining behavior.

chooses. The dotted line specifies an *information set* for Sharla. Basically it means that she does not know which of those two nodes she is at, since she did not observe Alice's choice. At the ends of the branches are the payoffs, again with player 1 listed first. There are ways to analyze extensive form games that are not possible in normal form games, but discussing these goes beyond the purview of this Element.

Now let us turn to evolutionary analyses. In evolutionary game theoretic models, groups of individuals are playing games, and evolving or learning behavior over time. *Dynamics* represent rules for how evolution or learning occurs. The question becomes: what behaviors will emerge over the course of an evolutionary process?

One central approach developed in biology to answer this question involves identifying what are called *evolutionarily stable strategies* (ESSs) of games. Intuitively, an ESS is a strategy that, if played by an entire population, is stable against mutation or invasion of other strategies.[8] Thus ESSs predict which stable evolutionary outcomes might emerge in an evolving population. Despite being used for evolutionary analyses, though, the ESS concept is actually a static rather than an explicitly dynamical one. Without specifying a particular dynamic, one can identify ESSs of a game.

Here is how an ESS is defined. Suppose we have strategies a and b, and let $u(a, b)$ refer to the payoff received for playing strategy a against strategy b. Strategy a is an ESS against strategy b whenever $u(a, a) > u(b, a)$, or else if $u(a, a) = u(b, a)$, then it is still an ESS if $u(a, b) > u(b, b)$. When these conditions hold, if an a type mutates into a b type, we should expect this new b type to die off because they get lower payoffs than as do. The strategies here are usually thought of as corresponding to fixed behavioral phenotypes because biological evolution is the usual target for ESS analysis. In the cultural realm, one can apply the ESS concept by thinking of mutation as corresponding to experimentation. The fit is arguably less good, but an ESS analysis can tell us what learned behaviors might die out because those who experiment with them will switch back to a more successful choice.

Let's clarify with an example. Suppose we have a population playing the game in Figure 2.1 – they all choose which coffee shop to meet at. Is a (all going to Starbucks) an ESS? The first condition holds whenever the payoff one gets for playing a against a is higher than the payoff for playing b against a. This is true. So we know a is an ESS by the first condition. Using the same reasoning, we can see that b (all going to Dunkin' Donuts) is an ESS too. Intuitively this

[8] This concept was introduced and defined by Maynard-Smith and Price (1973). Another related concept is the evolutionarily stable state, which need not involve only one strategy.

should make sense. When we have a population of people going to one coffee shop (who all like to coordinate) no one wants to switch to the other.

The second condition comes into play whenever the payoff of *a* against *a* is the same as the payoff of *b* against *a*. This would capture a scenario where a new, invading strategy does just as well against the current one as it does against itself. We can see that *a* will still be stable, though, if *b* does not do very well against itself.

As mentioned, Nash equilibria are important from an evolutionary standpoint. In particular, every ESS will be a Nash equilibrium (though the reverse is not true). So identifying Nash equilibria is the first step to finding ESSs.

As we will see in Section 8, philosophers of biology have sometimes been critical of the ESS concept. A central worry is that in many cases full dynamical analyses of evolutionary models reveal ESS analyses to be misleading. For example, some ESSs have very small *basins of attraction*. A basin of attraction for an evolutionary model specifies what proportion of population states will end up evolving to some equilibrium (more on this shortly). In this way, basins of attraction tell us something about how likely an equilibrium is to arise and just how stable it is to mutation and invasion. Another worry is that sometimes stable states evolve that are not ESSs, and are thus not predicted by ESS analyses.

We will return to these topics later. For now, let us discuss in more detail what a full dynamical analysis of an evolutionary game theoretic model might look like. First, one must make some assumptions about what sort of population is evolving. A typical assumption involves considering an uncountably infinite population. This may sound strange (there are no infinite populations of animals), but in many cases, infinite population models are good approximations to finite populations, and the math is easier. One can also consider finite populations of different sizes.

Second, one must specify the interactive structure of a population. The most common modeling assumption is that all individuals belong to a single population that freely mixes. This means that every individual meets every other individual with the same probability. This assumption, again, makes the math particularly easy, but it can be dropped. For instance, a modeler might need to look at models with multiple interacting populations. (This might capture the evolution of a mutualism between two species, for example.) Or it might be that individuals tend to choose partners with their own strategies. Or perhaps individuals are located in a spatial structure which increases their chances of meeting nearby actors.

Last, one must choose dynamics which, as mentioned, define how a population will change over time. The most widely used class of dynamics – *payoff* *monotonic dynamics* – makes variations on the following assumption: whatever strategies receive higher payoffs tend to become more prevalent, while strategies that receiver lower payoffs tend to die out. In particular, the *replicator dynamics* are the most commonly used dynamics in evolutionary game theory.[9] They assume that the degree to which a strategy over- or underperforms the average determines the degree to which it grows or shrinks. We can see how such an assumption might track evolution by natural selection – strategies that improve fitness lead to increased reproduction of offspring who tend to have the same strategies. This model can also represent cultural change via imitation of successful strategies and has thus been widely employed to model cultural evolution.[10] The assumption underlying this interpretation is that group members imitate strategies proportional to the success of these strategies. In this Element, we will mostly discuss work using the replicator dynamics, and other dynamics will be introduced as necessary.

Note that once we switch to a dynamical analysis, payoffs in the model no longer track utility, but instead track whatever it is that the dynamics correspond to. If we are looking at a biological interpretation, payoffs track fitness. If we are looking at cultural imitation, payoffs track whatever it is that causes imitation – perhaps material success.

Note also that evolutionary game theoretic dynamics do not usually explicitly model sexual reproduction. Instead, these models usually make what is called the "phenotypic gambit" by assuming asexually reproducing individuals whose offspring are direct behavioral clones. Again, this simplifies the math, and again, many such models are good enough approximations to provide information about real, sexually reproducing systems.

To get a better handle on all this, let us consider a population playing the coordination game. Assume it is an infinite population and that every individual freely mixes. Assume further that it updates via the replicator dynamics.

The standard replicator dynamics are *deterministic*, meaning that given a starting state of a population, they completely determine how it will evolve.[11] This determinism means that we can fully analyze this model by figuring out what each possible population state will evolve to. We can start this analysis by looking at the equation specified for this model and figuring out what

[9] These were introduced by Taylor and Jonker (1978) with the goal of providing a truly dynamic underpinning for ESS analyses.

[10] The replicator dynamics are the *mean field equations* of explicit models of cultural learning. This means that they represent the average, expected change in these stochastic models (Weibull, 1997).

[11] A *stochastic* dynamic on the other hand will incorporate some randomness so that, often, one population state has the potential to end up at different equilibria.