



University of Pennsylvania
ScholarlyCommons

Departmental Papers (Biology)

Department of Biology

7-2013

Biological Institutions: The Political Science of Animal Cooperation

Erol Akçay

University of Pennsylvania, eakcay@sas.upenn.edu


Joan Roughgarden

James Fearon

John Ferejohn

Barry R. Weingast

Follow this and additional works at: http://repository.upenn.edu/biology_papers

 Part of the [Animal Sciences Commons](#), [Behavior and Ethology Commons](#), [Biology Commons](#), [Genetics Commons](#), and the [Population Biology Commons](#)

Recommended Citation

Akçay, E., Roughgarden, J., Fearon, J., Ferejohn, J., & Weingast, B. R. (2013). Biological Institutions: The Political Science of Animal Cooperation. *Social Science Research Record*, 1-46. <http://dx.doi.org/10.2139/ssrn.2370952>

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/biology_papers/10
For more information, please contact repository@pobox.upenn.edu.

Biological Institutions: The Political Science of Animal Cooperation

Abstract

Social evolution is one of the most rapidly developing areas in evolutionary biology. A main theme is the emergence of cooperation among organisms, including the factors that impede cooperation. Although animal societies seem to have no formal institutions, such as courts or legislatures, we argue that biology presents many examples where an interaction can properly be thought of as an informal institution, meaning there are evolved norms and structure to the interaction that enable parties to reach mutually beneficial outcomes. These informal institutions are embedded in the natural history of the interaction, in factors such as where and when parties interact, how long and how close they stay together, and so on. Institutional theory thus widens the scope of behavioral ecology by considering not only why animals evolve to choose the strategies they choose, but also asking both why it is that they find themselves in those particular interaction setups and how these particular interactions can be sustained. Institutions frequently enable interacting parties avoid inefficient outcomes and support efficient exchange among agents with conflicting interests.

The main thesis of this paper is that the organization of many biological interactions can properly be understood as institutions that enable mutually beneficial outcomes to be achieved relative to an unstructured interaction. To do this, institutions resolve or regulate the conflicts of interests among parties. The way conflicts of interests affect the outcome depends on the structure of the interaction, which can create problems of commitment, coordination and private information. Institutional theory focuses on how to address each of these issues, typically focusing on the development of social norms, rules, and other constraints on individual behaviors. We illustrate our thesis with examples from cooperative breed and genes as within-body-mechanism-design.

Disciplines

Animal Sciences | Behavior and Ethology | Biology | Genetics | Population Biology

Biological Institutions: The Political Science of Animal Cooperation

Erol Akçay, Joan Roughgarden, James Fearon,
John Ferejohn, Barry R. Weingast, et al.¹

Version: July 2013

Social evolution is one of the most rapidly developing areas in evolutionary biology. The main theme in the evolution of social behavior is the emergence of cooperation between organisms, including the factors that impede cooperation. Political science studies cooperation and conflict, and the social structures these produce in the most socially complex animal, humans. We argue that both political science and evolutionary biology will benefit from more cross-disciplinary interaction, with each borrowing methods and perspectives from the other. In this paper, we focus on what political science can offer to biology in terms of concepts and methods. Specifically, we argue that a body of theory

¹ Department of Biology, University of Pennsylvania; Department of Biology, Stanford University; Department of Biology, Stanford University; Department of Political Science, Stanford University; Department of Political Science Stanford University and School of Law, NYU; Department of Political Science and Hoover Institution, Stanford University.

developed in political science that focuses on social, political, and economic institutions presents a potentially groundbreaking avenue for interdisciplinary synergy between political science and biology.

Our call for cross-disciplinary work between political science and biology is not without precedent: it was a paper by Robert Axelrod, a political scientist, and William Hamilton, an evolutionary biologist, that was seminal to a large part of the theoretical and empirical literature on the evolution of cooperation (Axelrod and Hamilton, 1981). More recently, Conradt and List (2009) initiated a project for interdisciplinary work on collective behavior in animal groups, a project that has already started bearing fruit. We follow these successful precedents and believe that the scope for interdisciplinary collaboration between political science and biology is wider than heretofore recognized. This paper presents some ideas that emerged in a meeting between political scientists and biologists.

Institutional theory and animal behavior

Human behavior does not take place in vacuum. Our actions are guided and constrained by social rules, norms and organizations. These rules, norms, and organizational structures -- collectively called institutions -- are not dictated from above by supernatural forces, but emerge in the course of history as a result of past decisions by individuals, groups and societies. Examples include the evolution of systems of government, the judicial system or the regulatory structures of economic institutions. Nor do institutions need to be written in laws and regulations; informal rules and conventions structure human interactions as much as laws, as anyone relocating to a new country can attest to. Institutional

theory in political science studies how institutions affect individual behavior and how and why they emerge and evolve over time (North, 1991, Ostrom, 1991). In doing so, it places individual behavior in the context of social organization, which consists of formal or informal rules and norms, and constrains individual behavior. The result is an interlocking system of social structure.

There are no demonstrated instances of formal institutions such as legislatures, courts and committees in animal societies. We argue, however, that biology presents many examples where an interaction can properly be thought of as an informal institution, meaning there are evolved norms and structure to the interaction that enable parties to reach mutually beneficial outcomes. These informal institutions are embedded in the natural history of the interaction, in factors such as where and when parties interact, how long and how close they stay together, and so on. In this sense, institutional theory widens the scope of behavioral ecology by considering not only why animals evolve to choose the strategies they choose, but also asking both why it is that they find themselves in those particular interaction setups and how these particular interactions can be sustained.

Given the ubiquity of institutions in human social and economic life, a first question becomes what purpose they serve. Institutions frequently enable interacting parties avoid inefficient outcomes and support efficient exchange among agents with conflicting interests. An inefficient outcome means that the gains from exchange or cooperation are not fully captured; or, more narrowly, that “money is being left on the table”: people fail to achieve outcomes that would

make everybody better off. The classic example of this is the prisoners' dilemma game, where regardless of the other's behavior, each prisoner has an incentive to defect rather than cooperate. These incentives for defecting preclude cooperation, which is the efficient outcome in a single-shot interaction.

However, the dichotomous setup of the prisoners' dilemma with a single efficient outcome belies the potential complexity of the problem of ensuring efficiency. More realistic games including the repeated prisoners' dilemma, have many different outcomes that are efficient. To give a biological example, consider the case of two predators sharing a prey killed. An outcome is efficient if all the meat gets eaten. Killing the prey requires some measure of cooperation between the two predators, but they still have conflicting interests between different efficient outcomes – who eats how much – and their conflicting interests may lead them to fail to cooperate. For instance, if the weaker of the two predators expects to be excluded from the kill after the fact, it might fail to cooperate to bring it about in the first place. Humans face such problems constantly, and an incredibly diverse range of social institutions have arisen in myriad contexts to enable the human equivalent of the stronger predator in this example to commit itself to share the kill. Thus while cooperation is regularly needed to achieve efficient outcomes, conflicts of interests can lead to inefficient levels of cooperation.

The main thesis of this paper is that the organization of many biological interactions can properly be understood as institutions that enable mutually beneficial outcomes to be achieved relative to an unstructured interaction. To do

this, institutions resolve or regulate the conflicts of interests between parties. The way conflicts of interests affect the outcome depends on the structure of the interaction, which can create problems of commitment, coordination and private information. Institutional theory focuses on how to address each of these issues, typically focusing on the development of social norms, rules, and other constraints on individual behaviors. We illustrate this with an example below.

Commitment, Coordination, and Private Information

Commitment: The merchant guild

Medieval Europe was a tough place to do business, especially over long distances (Milgrom et al., 1990, Greif et al., 1994). The rule of law was by no means assured, and merchants' property rights and the enforcement of contracts were not guaranteed by agreements between different polities as they are in today's world. Greif et al. (1994) ask how in such an environment long-distance trade could emerge. They consider a city whose ruler benefits from trade within his city by taxing it, but also who is capable of robbing any single trader that comes to his city. Greif et al. show that, because the ruler cannot commit not to rob any individual merchant, such a city cannot sustain trade at levels that maximize the total level of trade and exchange as well as the profits for merchants and the ruler combined (i.e. efficient levels). This result occurs because at the level of trade that maximizes total profits, the value of business with any single merchant is marginal. This means that even if the cheated merchant retaliated and never traded with the city again, the ruler's loss would

not be great, hence he would do better by robbing the merchant. In this case, commitment problems preclude efficiency.

Greif et al. argue that mediaeval merchants solved this problem by organizing themselves into merchant guilds. Once a merchant guild is in place, it can declare a ban on trade with a ruler that cheated one of its members. Thus, the ruler would face retaliation from not a single merchant with marginal value to the ruler, but from the whole guild, whose value to the ruler would be substantial. It becomes in the ruler's own interest not to cheat *any* merchant belonging to the guild; consequently, merchants belonging to the guild can trust his promise not to do so, solving the commitment problem. The guild has to be sufficiently large to create the right incentive for the ruler to honor his commitment not to expropriate wealth from traders. And of course the members have to have incentives to honor their own commitments to obey the ban on trade.

Two features of this model are worth noting: first, in this model, the merchant guild does not form to demand a better deal (e.g. a lower tax rate) from the ruler (although subsequently, it can do that as well). Rather, the merchant guild enables the merchants to coordinate in the face of transgressions, thereby allowing the ruler to commit to honor whatever deal was struck. The guild institutions thereby increase the payoff to both the merchants and rulers. Second, in solving the commitment problem for the ruler, the merchant guild creates another one, namely that of coordinating merchants and getting them to commit to honor bans imposed by the guild. Greif et al. (1994) argue that merchants within guilds had both complex institutions and complex interrelationships with

each other that created incentives for them to honor the decisions made by the guild.

A related model deals with the emergence of the Law Merchant to resolve trade disputes in the Mediaeval age (Milgrom et al., 1990). In this model, a given pair of traders interacts only once with each other, but has the option of reporting transgressions to a third party (the Law Merchant), who keeps records of transgressions that are not remedied. At the equilibrium, all traders consult (by paying a fee) the Law Merchant before trade about whether their prospective partners have outstanding judgments against them, withhold cooperation from those who do, and report any transgression to the Law Merchant after the interaction. This equilibrium sustains cooperation, because a trader who cheats a partner loses all future business, even though he will never interact with that particular partner again. Here again, the institution of the Law Merchant resolves the commitment problem faced by traders by putting in place appropriate incentives.

Commitment in a biological institution: the cleaner fish

We now illustrate how institutional theory can be used to gain insight on a biological system, using the cleaner fish mutualisms as our example. Cleaner fish are species that inspect other fish for ectoparasites and remove these. The best-studied examples are the cleaner wrasses *Labroides dimidiatus* and *L. phthyrophagus* that live in coral reefs throughout the Indian and Pacific Oceans. These fish occupy small territories, called cleaning stations, and are visited by other fish (clients) that they inspect and clean. In exchange for their service,

cleaners get to eat the ectoparasites, but they can also feed on the healthy mucus and scales of their clients (thereby hurting them), and prefer this to the ectoparasites (Bshary and Grutter, 2002).

The question then becomes, how clients keep cleaners from cheating by feeding on healthy tissue. Different species of clients have different options available to them: a few of the client species are predatory and can simply eat a cheater while others exercise choice between different cleaners (Bshary and Schäffer, 2002)². Yet another class resorts to a punishment strategy, chasing the cleaner around after being cheated (Bshary and Grutter, 2002). Clients are also known to observe other clients interacting with their prospective cleaner (called image-scoring, Bshary, 2002), and remarkably, cleaners seem to be able respond to novel behavioral patterns exhibited by clients and adjust their behavior in order to optimize their gains (Bshary and Grutter, 2005, Bshary and Grutter, 2006).

These findings demonstrate that the cleaner fish system exhibits complex social organization geared towards enabling mutually beneficial exchanges, much like human economic activity. In fact, the idea that systems such as this are governed by market forces similar to economic life has been advanced previously (Noë and Hammerstein, 1995, Noë et al., 2001). Market theory relies on implicit assumptions about how the market operates, such as the costless and

² One might also ask how a predator commits not to eat the cleaner once the cleaning is almost done. The most plausible answer, as suggested by Trivers (Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.) is that the benefit of repeatedly being cleaned is higher than the one-time benefit of eating a small fish. In fact, predators generally seem to have reduced aggressiveness toward all fish when at the cleaning station (Cheney, K. L., Bshary, R. & Grutter, A. S. (2008) Cleaner fish cause predators to reduce aggression toward bystanders at cleaning stations. *Behavioral Ecology*, 19, 1063-1067.).

effective enforcement of contracts and the availability of information about prices (North, 1991); this approach fails to explain how efficient enforcement of contracts is sustained. In contrast – and as illustrated in the merchant guild example – institutional theory studies how institutions can provide these features to the marketplace. In the cleaner fish case, for example, enforcement of contracts means that a cleaner must signal and commit to a “price” for its services (e.g. how much healthy tissue it will consume per minute of cleaning). That this cannot be automatically assumed is demonstrated by a subset of cleaners who signal clients their cooperativeness, but then go on to feed on healthy tissue (Bshary, 2002). Market choices can prevent such transgressions under some circumstances. But as in the merchant guild example, when cleaners are saturated and the cleaning interactions happen at efficient rates, the value of any single client to the cleaner will be marginal, and a commitment problem presents itself. Furthermore, a client fish that interacts with multiple cleaners would be limited to information that it can gather either by directly interacting with a cleaner or by observing a cleaner interacting with another client. Obtaining information in this way is likely to be costly and easily manipulated. Both of these problems would be solved if all clients had access to a long-term repository of the cleaner’s performance. Candidates for storing such information are the territorial clients with access to a single cleaner. If a cleaner cheats one of its prior clients, its territorial clients can make this known to all future clients, which would then do best to avoid the cheating cleaner, similar to the Law Merchant model of Milgrom et al. (Milgrom et al., 1990). In this way, the cost of cheating for the cleaner is

raised from the marginal cost of losing one client to the cost of losing many or all clients; this change, in turn, makes it in the cleaner's best interest to not cheat and therefore solves the commitment problem. Such an institution also solves the cooperation problem between the territorial clients and cleaners, since now cleaners have incentives to be cooperative towards these clients. Overall, this argument predicts that the interactions between territorial clients and choosy clients should play an important role in maintaining cooperation in the cleaner fish system.

This example shows how the theory of institutions can be applied to animal behavior to generate new empirical and theoretical research questions. In the next section, we provide a brief survey of a few of the major questions in institutional theory and the approaches used to solve them, also pointing out their connections with existing biological theory. We then discuss in some detail one of the major theoretical tools of institutional theory, mechanism design, and two ways it can be applied to biology. We follow this by suggesting some additional biological phenomena that seem most profitable for applying an institutional approach, and general theoretical problems that need to be considered.

Coordination: standard setting

Coordination problems in social interactions occur when there are multiple viable courses of action, but their benefits are only realized when the interacting individuals can agree on a given course. Such situations arise in diverse settings, ranging from competition of two different high-capacity optical drive formats to the movement decisions of an elephant herd. Coordination failure can be a major

factor precluding efficiency, even in cases where the interests of the parties are largely concordant.

In an influential paper, Farrell and Saloner (Farrell and Saloner, 1988) investigate whether the coordination problem between two players is best resolved through a committee where players bargain, or through the “open market” where both players come forward with their own actions and hope that their opponent follows suit. The tradeoff between the two institutional structures is that the committee ensures coordination, but imposes negotiation costs (in particular, delays in agreement) while the market minimizes delay costs while creates the hazard of mis-coordination if both players commit to different actions at the same time. Farrell and Saloner show that a hybrid institution that prescribes bargaining in a committee while also allowing players to individually commit to a course of action at every possible stage does best compared to both of the pure institutions. More recently, Farrell and Simcoe (Farrell and Simcoe, 2009) study the optimality of war-of-attrition type rules for deciding on industry standards when two proponents have private information about the quality of their proposals. They find that when there is no vested interest (i.e. no conflict over the eventual standard), the war-of-attrition game chooses the best standard without delay, and thus achieves the first-best outcome. However, when the vested interest is high enough, it becomes optimal to employ an institution that allows the war-of-attrition to proceed until a specified time and if the game is unresolved at that time, chooses an outcome randomly. Both players prefer such

an institution to the unchecked war-of-attrition before they know their proposals' quality, but it might not be preferred during the war-of-attrition.

The biological counter-part to this coordination problem can be found in collective decision-making with multiple alternatives (Conradt and Roper, 2005). The results of Farrell and Saloner (Farrell and Saloner, 1988) suggest that a combination of consensus building through communication and individual initiative will be optimal for cases with perfect information about the alternatives. This would predict that we should observe a mix of consensus decisions and individual (despotic) decisions even in cases where there is no reason to expect one particular individual to be the decision maker. On the other hand, Farrell and Simcoe's (Farrell and Simcoe, 2009) result suggest that when the decision is held-up between two parties with conflicting interests (e.g. seeking water vs. seeking food), it might be optimal for a third, uninterested party, to make the decision in a random way. Such a mechanism can manifest itself as the individual in best condition in the group (e.g. most satiated and hydrated) making arbitrary decisions, since such an individual would represent the closest approximation to a neutral party in the group. The issue of coordination is also intimately related to the issue of revealing private information.

Private information: conflicts and information aggregation

Private information in social interactions can hinder efficient outcomes because parties that have information will usually have incentives to not reveal it truthfully. For example, in the situation modeled by Farrell and Simcoe (Farrell and Simcoe, 2009), individuals have no intrinsic incentives to reveal the quality of

their proposals truthfully; such incentives are instead supplied by having the players “pay” through the costly war-of-attrition stage. Ensuring that private information is truthfully revealed can even be a problem even in cases of completely concordant interest; see “*Voting and information aggregation*” below. The problem of private information crops up in many different areas of political science and has also attracted a great deal of attention in biology, most prominently in costly signaling theory and war-of-attrition games. We consider these in more detail below.

Major questions in institutional theory

Contemporary institutional theory deals with questions that range from the structure and functioning of the legislative process (Baron and Ferejohn, 1989), to economic development (North, 1991), the evolution of cooperation (Bowles 2006), and the management of common resource pools (Ostrom, 1991). We do not attempt to give a complete overview of institutional theory in this section. Rather, we focus on some papers from three major fields to illustrate the approaches taken in the field and how they could be applied to biology: international conflict, voting and information aggregation in groups, and the theory of the firm.

International conflict and how to prevent it

Political scientists have a long-standing interest in violent political conflicts, for many obvious reasons. In recent years a common approach has been to start from the observation that since wars destroy valued resources and often

pose significant risks to the political leaders who start them, they are inefficient. Given any outcome of the war, both parties should prefer a peaceful resolution with those terms to having fought a war and settled at the exact same terms. So why do wars and other costly violent political conflicts sometimes occur?

One answer is that private information of each state about its own attributes, such as military capability or value it places upon the disputed territory, will preclude finding a mutually acceptable peaceful settlement (Fearon, 1995). When two states are engaged in pre-war negotiations, both can have interest in withholding or misrepresenting their private information, either to get a better deal if they settle, or because the other party cannot commit to honor the settlement and not attack given the disclosed information. Therefore, with each state exaggerating its strengths and value it places upon the object of contention, no feasible negotiated outcome may exist that looks preferable to both states *ex ante*. War may then follow as a means of credibly revealing (or bluffing about) one's private information.

The implication from this argument is then that to prevent wars, institutions need to be set up that give states incentives to credibly reveal their private information without actually fighting. Common practices in international "militarized disputes" that are short of a full-blown war – such as mobilizing troops and issuing public statements refusing to back down – may be seen as institutions that enable and structure the interpretation of "costly signaling." Recent literature on the political science of war deals with what these institutions can achieve and when wars can be prevented (e.g. Fey and Ramsay, 2009,

Meirowitz and Sartori, 2008). One result from this literature is that when states' private information is correlated (as for example would occur with private information about military capabilities) then regardless of the details of the negotiation process between the states there must be a positive risk of fighting if total costs for war fall below a threshold value. More recently, Meirowitz and Ramsay (2009) extend these results to situations where states react to different international institutions by altering their military capacity, and showed that the probability of war given a certain arming level is independent of the institutional structure for negotiations. These strong results are obtained using a powerful theorem of mechanism design, called the revelation principle (Myerson, 1979, see also the section on mechanism design below). Finally, a model by Slantchev (2003) suggests that states can rationally go to war even when they have complete information about each other's military capacities, etc., as a means to secure more favorable settlements. In this case, war can be seen as a costly mechanism to force coordination on a particular settlement.

Biologists have long used the metaphor of war and peace for agonistic behavior between animals (Maynard Smith and Price, 1973), and some aspects of previous theory bear resemblance to the political science literature on the subject. One example is found in models of signaling before agonistic interactions: Enquist (1985) shows that in situations with two animals that have private information about their valuation of the resource or their fighting ability, stable signals can evolve that are informative and allow animals to avoid fighting some of the time. This is closely parallel to the literature on "crisis bargaining" in

International Relations. On the other hand, Kim (1995) models a symmetric game where individuals do not have private information, and shows that conventional signals with costs can evolve so as to lower the risk that costly conflicts will occur due to coordination failure.

This focus in biology on how animals avoid fighting costs parallels the political science literature on costly signaling and war. However, biological models tend to be restricted to particular game setups, and usually do not allow negotiated partitions of the resource that is being fought over (i.e. multiple possible bargains) as a result of the pre-fighting interactions. Bargaining models commonly used in the theory of social and political institutions can be used to extend existing biological theory to cases where a division of the resource is plausible, such as territorial interactions or bargaining over resource exchange (Pereira et al., 2003, Akçay and Roughgarden, 2007b). Furthermore, a mechanism design approach can help to extend these results to generate “game-free” results about conflict, as in Fey and Ramsay (Fey and Ramsay, 2009), or results about the evolution of armaments, similar to states’ arming decisions in Meirowitz and Ramsay (2009).

Voting and information aggregation

Voting and information aggregation are questions at the very heart of political science, and theoretical work on these issues goes back to the 18th century. This area has recently begun to see cross-disciplinary collaboration between biologists and political scientists (List, 2004, Conradt and List, 2009). Voting theory is concerned, among other things, with how and whether efficient

outcomes can be achieved when multiple individuals with potentially different interests and private information have to make a collective decision. This literature usually takes the point of view of a “social designer”, whose goal is to satisfy group-level criteria for the aggregate decisions. As pointed out by Conradt and List (2009), this contrasts with the biologists’ approach, which commonly looks for solutions that are optimal from the point of view of single individual. Thus, biologists have modeled optimal decision-making rules based on their effects on individuals’ fitness, which in general will be not aligned (Conradt and Roper, 2009).

Interesting problems arise when individuals involved in joint decision-making have private information about their own conditions or the environment. If individuals have a common interest but have an independent estimates of what action best leads to the common interest, and if all have the same and better-than-even-chance of being accurate, then having more individuals reveal their information and taking the average of all estimates, on average, should improve the accuracy of the decision (known as the Condorcet Jury Theorem). Some complications arise when individuals are not equally likely to have accurate estimates, or if individual’s estimates are correlated (see the examples in Conradt and List, 2009), although the result holds approximately for small departures.

More serious problems arise, however, when group members decide strategically about whether or not they will reveal their information, or vote strategically. Surprisingly, in such cases it may pay for individuals to withhold or

misrepresent their information, even when there is complete concordance of interest within the group (Austen-Smith and Banks, 1996, Austen-Smith and Feddersen, 2009). The reason is that an individual affects the vote outcome only if he or she is pivotal meaning that the focal individual is indifferent about how she votes in all other cases. But the event that one is actually pivotal implies that the other individuals are voting in a particular way (e.g. all are voting “Yea” in a setting requiring unanimity), which allows the focal individual to update her beliefs about the state of the world. In such cases, voting against one’s private signal can be optimal, and therefore votes might cease to be informative about the private signals. In general, the incentives to misrepresent one’s information are determined by how likely it is that a player will cast the decisive vote in determining the outcome. Thus, these incentives are a bigger problem in smaller groups, since each individual has a higher probability of being pivotal, and present lesser problems in larger groups, where the Condorcet Jury Theorem approximately survives strategic information sharing and voting (Feddersen and Pesendorfer, 1997).

Not surprisingly, the problem of strategic voting and misrepresentation is aggravated when there is real conflict of interests within the group over the relevant decision. More recent work uses a mechanism design approach to this problem and studies incentives for truth telling in collective decision making. For example, Meirowitz (2006) shows that outside transfers to individuals as a function of their revealed information can create incentives for truthfully representing their information, and that the magnitude of the required transfers

becomes smaller as group size gets larger (due to each individual having smaller chance of being pivotal). These results have not yet seen use in the biology of group behavior, but have connections to the costly signaling theory in biology, where signal costs can be interpreted as negative transfers (see below).

The theory of the firm

The theory of the firm is a part of the institutional theory that originated in economics. A firm in economics is an organization that produces goods and services outside the marketplace, by means of contracts that last much longer than each action the agents take (e.g. producing a single unit of goods). These contracts frequently concentrate the means of production and decision-making in some agents and remunerate others in return. The first question in the theory of the firm is why firms exist at all, i.e. why do agents organize themselves into long-term relationships regulated by contracts as opposed to finding each other in the marketplace and achieving production by on-the-spot transactions? This question, first posed by Coase in his influential essay (Coase, 1937), has stimulated a large body of research in law, economics, and political science.

Coase himself proposed that firms form to reduce the costs that arise from making repeated transactions in the marketplace, such as the cost of finding out the market price of goods and services and negotiating over terms of agreement. The transaction costs theory has further been extended to other causes of costs. One example is specificity of production assets to each other (e.g. a supplier building a plant next to a manufacturer's factory), which removes the possibility of partner choice in the open market. In such cases, situations that are not covered

in the original agreement between parties or that arise from unexpected events create incentives for the advantaged party to try to appropriate gains (or avoid losses) from those situations, to the detriment of the other party – much as in the example sketched above of a strong and weak predator possibly cooperating to hunt prey. In the absence of vertical integration (i.e. the integration of a supplier firm with a firm who purchases from it) can lead to underinvestment relative to what is efficient, because both parties expect that their sunk costs from the specific investments will be appropriated by the other. Klein et al. (Klein et al., 1978) argue that efficient transactions can be achieved in such cases when one party owns both assets, instead of relying on repeated market transactions.

Following up on this argument, Grossman and Hart (1986) formally modeled the incentives to parties to make such relationship-specific investments with incomplete contracts, i.e. contracts that do not account for all possible contingencies that might happen in the future. In this setting, vertical integration of the production assets can solve the problem of underinvestment, by assigning control of both assets to one party (the firm) reduces the incentive to appropriate sunk costs after the investments are made and hence optimal investment decisions will be taken beforehand. In the absence of vertical integration, appropriation of assets transfers resources from one firm to another; whereas vertical integration means the transfer occurs within the same firm.

On a related vein, Williamson (Williamson, 1979) argues that long-term repeated transactions with highly specific or idiosyncratic assets should be governed by what he calls relational contracts, which specifies the roles of two

parties in an ongoing relationship, rather than particular actions they need to take in each possible state of the world. Relational contracts can take the form of partnerships, or an employee-employer relationship.

How can these ideas be applied in biology? Consider the striking contrast between birds, where the overwhelming majority of species exhibit social monogamy (i.e. a male and a female raising offspring together), versus mammals, where the overwhelming majority of species is polygamous. Previous thinking on this pattern mostly departs from the assumption that females prefer monogamy, and males polygamy (Clutton-Brock, 1991). The question is then when females can impose monogamy on males, either directly by choosiness, or indirectly through their distributions in space (Clutton-Brock, 1989).

Viewing offspring rearing as analogous to the production of goods by a firm leads to a new perspective. In mammals, lactation implies that females control offspring provisioning, so that males cannot directly invest into that component of care. However, males can invest indirectly into provisioning through feeding the female (or allowing her to forage undisturbed), and also into other components of care, such as predator protection. Moreover, in most mammals, females are mobile while gestating and offspring are either mobile shortly after birth, or can be carried around. These features of mammalian breeding biology mean that females can receive help from different males (e.g. by moving between territories) without necessarily losing their offspring (especially if they have copulated with multiple males, Wolff and Macdonald, 2004). A female bird in an altricial species, however, cannot move her eggs or

nestlings to another nest without killing them; likewise, the male cannot share the nest with another female while another brood is in it. Hence, investments by the male and female into breeding are idiosyncratic in most bird species, whereas they are far less relationship-specific in mammals. In these different settings, the theory of the firm predicts that interactions between mates in birds tend to be governed by “relational contracts”, i.e. longer-term commitments such as the pair-bond. On the other hand, interactions between mates in mammals can be maintained by repeated shorter-term commitments (or on-the-spot transactions) since each party maintains an outside option due to their less partner-specific investment. Notice that this argument runs counter the traditional view of the contrast between the two mating systems, as illustrated by Clutton-Brock (1989), who posits that monogamy is expected if it is impossible or very costly for one male to monopolize more than one female, either because females choose against already mated males (as in birds), or are distributed so sparsely that it is impractical for males to maintain a territory with two females in it. Our argument is closer (with some differences) to the position advanced by Roughgarden (2009) that male promiscuity in mammals reflects a counter strategy to the female’s control of the offspring.

This hypothesis can be tested by taking advantage of the variation in offspring mobility in different mammal and bird species: our argument predicts that species with more mobile females and offspring will be more likely to have social polygamy. Conversely, species where breeding females are more

localized to exclusive locations offspring are less mobile will be more likely to exhibit social monogamy.

The theory of the firm will also be useful in understanding the organization of mutualisms. For example, one can ask why nitrogen fixing rhizobia are housed in plant organs called nodules that represent high initial and specific investment by the plant, while plant interactions with mycorrhizae do not have such partner-specific structures.

Mechanism design

We now turn to one of the most important and powerful tools of institutional theory, mechanism design. Most of the work in game theory specifies a game exogenously, and predicts outcomes supposing self-interested agents with some level of computational capacity and access to certain public and private information. Mechanism design inverts this approach: it specifies the information structure and some range of games, and looks at which games produce outcomes with desired properties, such as efficiency or maximizing “the principal’s” expected payoff. Thus, a mechanism in game theory describes the different rules of the game that may structure an interaction between two or more individuals. Note that this usage of the term “mechanism” is different than the term’s meaning in biology, where it refers the processes that bring about biological phenomena. In the case of behavior, these biological mechanisms can also be called “proximate causes”. For example, the mechanism for a behavior such as aggression might involve specific neural circuits in the brain or the testosterone level in the body. In the remainder of this paper, we use the term

“mechanism” in the game theoretic sense, and refer to biological mechanisms as “proximate causes”, to avoid possible confusion. The “rules of the game” in a social interaction between animals are ultimately a function of the proximate causes of behavior, hence the two meanings of the term “mechanism” are closely related. In fact, as we argue below, mechanism design can be used to understand the organization of proximate causes of behavior.

Some mechanism design models take up the perspective of one of the parties in the interaction who has power to alter the game structure (such as the parent company in a conglomerate Groves, 1973), while others adopt a more disinterested “social designer” perspective (for example, authors of a constitution). In biology, the former approach may be appropriate in social interactions where there is an actor (e.g. a parent bird at a nest) who has control over the setting that others (e.g. the nestlings) interact in. Recently, Roughgarden and Song (in preparation) has modeled parental provisioning strategies using Groves’ model of incentive compatible mechanisms, where the parent determines the payment scheme to the chicks as a function of their signals. Their model illustrates one way the mechanism design methodology can be used in a biological setting.

When used from a social planner perspective, mechanism design is a tool to characterize efficient outcomes that can be achieved by resolving conflicts of interests between individuals. We propose that this tool can be used to ask whether animal social systems approximate efficient mechanisms for dealing with the underlying strategic problem. This would be analogous to using the concept

of an evolutionarily stable strategy (ESS) to provide adaptive hypotheses for some behavior observed in nature. The main difference is that mechanism design would apply to how the interaction among individuals is organized and what effects this organization has on the joint outcome, instead of how single individuals behave. On the other hand, similar to the ESS approach, mechanism design would not completely answer the question of how these outcomes are actually achieved, which requires specification of the details of the strategies in a game, and leaves open the dynamical (and population-genetical) process of reaching such outcomes.

A foundational result of mechanism design theory is the revelation principle (Myerson, 1979), which states that any Bayesian equilibrium of any game with incomplete information can be represented by a special class of games, called direct mechanisms. In a direct mechanism, players simply announce their type (their private information) to a central arbiter, who then assigns payoffs (possibly including transfers) as a known function of all possible announcements (that is the mechanism). In an “incentive compatible” mechanism, payoffs structured as to make it optimal for players to reveal their information truthfully. The revelation principle implies that if an outcome can be implemented in a game, it can be implemented by a direct mechanism. This theorem allows considering only direct mechanisms when trying to find out when some outcome is achievable, the magnitude of transfers and costs that are needed to achieve an efficient outcome, or to characterize the best feasible

outcomes. Further, using the revelation principle, one can derive comparative statics results that are fairly independent of the details of the game.

The mechanism design approach has been used to show that in a number of important social settings where private information is relevant to the determination of optimal policies – for example, all manner of bargaining problems, or the design of a tax system – it is impossible to design an institution (mechanism) that will achieve first-best outcomes (unless there is some outside source of payments to players). Private information may pose an ineluctable cost when there are conflicting interests. The difference in payoffs between the first-best outcome and the so-called second-best outcome that is possible to achieve in an incentive compatible fashion is often termed “agency loss”.

Mechanism design in biology: costly signaling

A special case of incentive compatible mechanisms can be found in an hypothesis that is familiar to most behavioral ecologists: the handicap principle posits that individuals with conflicting interest and private information can communicate with each other honestly, provided that signals are costly (Zahavi, 1975, Grafen, 1990), and the cost function has a particular form (Grafen, 1990). The cost function here functions in the same way that a transfer in a mechanism design model would function. By imposing negative costs as a function of the message sent, the handicap principle ensures that it is individually optimal to signal truthfully; hence, costly signals represent incentive compatible mechanisms.

The two settings in which costly signaling has been most important are signaling of mate quality (Grafen, 1990) and parent-offspring communication about need (Godfray, 1991, Godfray and Johnstone, 2000). The main difference between these two types of models is that in models of mate quality signaling, the signal costs are condition-dependent (higher quality individuals pay lower costs), while offspring signaling models assume that costs are not condition-dependent, but benefits are (higher need individuals gain more from the same amount of food).

In one sense, signal costs in these models are wasted; they do not translate into fitness to either party. They represent agency loss, or a second-best outcome, relative to what could be achieved if, for example, the females could tell apart males of different quality without any costs; but such a system would not be incentive compatible. In fact, the agency loss in a costly signaling system can be so high as to render one or both parties worse off relative to no signaling (Rodríguez-Gironés et al., 1996). This happens when variation in the quantity to be signaled is restricted, and the benefits from signaling is low relative to the costs that ensure incentive compatibility (Godfray and Johnstone, 2000). The reason for the discrepancy between these two is that the former is determined by the average costs and benefits from signaling while the latter is dictated by the marginal costs and benefits.

Given that costly signaling can be understood as a mechanism in the game theoretic sense, we can apply the tools of mechanism design theory to it. Most signaling models deal with particular signaling setups, usually involving two

or few individuals (e.g. one parent and one or two offspring). The reason is that the complexity of the models increases rapidly when many individuals interact, with each other, due to the need to take into account all individuals' responses to each other. In these cases, the revelation principle can offer significant simplification: since any equilibrium of any game structure can be represented by a suitable direct mechanism where each individual only interacts with an (imagined) central arbiter, one can analyze a much simpler game for the purposes of making prediction on the outcome of the interaction, without considering the details of the behavioral game. This is especially important when multiple parties exchange information with each other and all react to everybody else's signals (as in a multi-offspring brood), or when costs of signaling are not automatic (such as energy costs of begging, or growing a large tail), but imposed socially through other individuals' actions (Clutton-Brock and Parker, 1995, Lachmann et al., 2001). These situations bear some resemblance to Meirowitz' model of collective decision making with communication and conflicting interests (Meirowitz, 2006). A mechanism design approach provides a powerful tool for advancing comparative hypothesis and empirical tests, and can be used to complement one based on empirically motivated models of interactions.

Evolutionary mechanism design theory

Thus, one can use mechanism design as a tool to characterize possible outcomes that can be achieved in an incentive compatible way, and get comparative statics results. Using mechanism design in this way is similar to using an ESS analysis to find strategies that are likely to be the result of long-

term evolution. On the other hand, mechanism design by itself does not tell us how the outcomes will be achieved. In other words, it doesn't answer what kind of game structure, or institution, needs to be in place so that the individuals will actually reach those outcomes. This question, of course, is one of the primary concerns of biologists studying behavior and its answer depends on the immediate determinants of individual behavior in a social interaction, i.e. the proximate causes of behavior.

The standard model of proximate causation in economics is the so-called "rational actor model": agents have beliefs and expectations over the state of the world and their partner's types, actions, etc., and can carry out complex calculations to find actions that maximize their utility (however defined) given these beliefs and act accordingly. Such a model underlies widely used solution concepts such as the Nash equilibrium and perfect Bayesian equilibrium. Biologists, on the other hand, have traditionally tended to assume implicitly an almost trivial model of proximate causation, where the behavior of an individual is determined by a single locus in a haploid genetic model. Recently however, both fields have been moving away from their respective models. Economists increasingly adopting models that take into account bounded rationality and cultural influences on preferences (Gintis, 2007), while biologists have been developing game theory models with explicit mechanisms of proximate causation (McNamara et al., 1999, Roughgarden et al., 2006, Akçay et al., 2009, Roughgarden, 2009).

Among the few biological models of proximate causation, the model of goal-orientated behavior by Akçay et al. (2009) provides many immediate linkages to the game theoretic literature. In this model, individuals have a genetically encoded objective function that represents the internal reward sensation of the agents and might be different than their material payoffs. Individuals act myopically to maximize their objective functions, which in effect assume the role of the utility function in a bounded-rational actor model. Accordingly, the behavioral dynamics of Akçay et al. result in a pure strategy Nash equilibrium of a game defined by the objective functions. The material payoffs to the individuals depend on their equilibrium actions and these payoffs in turn determine individuals' fitness. Thus, objective functions are under selection pressure through the fitness they induce. Applying this framework to a continuous prisoners' dilemma, Akçay et al find that objectives that place value on one's opponent's payoff can be evolutionarily stable. This model is mathematically equivalent to "indirect evolution" models in economics (Güth, 1995, Dekel et al., 2007), with perfect information about agents' types, because during the behavioral dynamics in Akçay et al., individuals effectively end up learning each other's objective functions.

Such non-selfish objectives represent an internal commitment to cooperate, even when this is against the actor's material interest in the short term (Güth and Kliemt, 2000). This contrasts with the way commitment is achieved in Greif et al's model of the Merchant Guild (Greif et al., 1994) or the proposed institution in the cleaner fish system, both of which rely on pure self-interest.

These two different modes of commitment require overcoming different challenges: the Merchant Guild has to be able to make the threat of collective retaliation by the Guild credible, while the other-regarding individual has to be sufficiently certain of the preferences of its partner in order to be not taken advantage of (Ok and Vega-Redondo, 2001). These cases illustrate that the same outcome (cooperation) can be implemented through different methods. Further, these two methods can operate at the same time. As an example, a strong argument has been made that human cooperation is maintained through punishment of cheaters, which makes cheating materially unprofitable. Yet punishing might also be materially costly, and hence Gintis and colleagues (Gintis et al., 2003, Gintis, 2003) argue that commitment to punishing is achieved through the evolution of other-regarding preferences. This shows that the proximate cause implementing the outcome of cooperation can involve both material incentives or threats, and internal commitments. We believe that empirically motivated models of proximate causation need to be used to complement methods mechanism design by asking how incentive compatible outcomes can be implemented through internal mechanisms and the structure of the interaction.

Another shortcoming of traditional mechanism design theory is that it is mostly silent about how the mechanism will come to be in place, i.e. how the rules of the game evolve. Similarly, evolutionary game theory models in biology virtually always assume that games are fixed as properties of the physical environment and do not consider evolution of the games themselves at all. Thus,

there is a need to develop an “evolutionary mechanism design” theory where rules of the games, or institutions are not imposed by an outside designer by fiat, but evolve through their fitness consequences for the participating individuals. This is still an active question in institutional theory (Greif and Laitin, 2004), and has an immense potential for synergism between biologists and political scientists.

Two recent models in biological literature have tackled the question of how a game might evolve is through individual selection (Worden and Levin, 2007, Akçay and Roughgarden, in preparation). Worden and Levin concentrate on how novel strategies with different payoff consequences can invade a population of players playing a Prisoner’s Dilemma and find that this process leads away from a Prisoners’ Dilemma towards a mutualism game. Similarly, Akçay and Roughgarden develop a population-genetic framework to complement the behavioral dynamics of Akçay et al. discussed above, and derive conditions that allow traits that provide incentives to cooperate in a prisoners’ dilemma game to invade and fix in a population. One interesting result from their analysis is that such an evolutionary process will frequently lead to genetic polymorphisms in the types of games played. This means that even with a single social interaction, multiple games with strategic properties ranging from conflict (as in the prisoners’ dilemma) to coordination (as in the battle-of-the-sexes) can co-exist in a population.

Institutions and levels of selection

The main themes of this paper is that institutions can organize interactions in ways that make self-interested parties achieve socially desirable outcomes, such as maximizing the productivity or efficiency of an interaction and that many animal social interactions may be organized in similar ways. An institution is necessarily a theoretical abstraction, but it must correspond to an observable phenomenon in the real world. In biology, the place to look for the institution is the natural history of interactions, i.e. the description of who interacts with whom, when, how, and under which conditions. Thus, the institutional perspective shifts attention from considering individual behaviors by themselves to considering the evolution of the whole social system the interaction is taking place in.

An important implication of the institutional perspective is that the organization of biological interactions might allow individuals to achieve outcomes that maximize efficiency or productivity at an aggregate level, even though each party is experiencing selection individually. This implication puts a new twist to an ancient debate in evolutionary biology between those who favor group selection as a powerful cause of evolution and argue that group-level adaptations can be observed in nature, and those who argue that selection exclusively acts at the individual (or genic) level (for an overview, see Okasha, 2006).

The institutional perspective suggests a different approach to the question of whether evolution can lead to optimization at the level of an aggregate system; say a coral reef with its entire set of interacting species including the cleaner wrasse and its clients. Neither kin selection nor group selection can act on a

coral reef ecosystem, since species are unrelated to each other and there is no population of coral reefs that are subject to differential mortality and reproduction as a whole. However, the institutional perspective raises the possibility that selection acting separately within different species at the level of the individual can lead to maximization of efficiency at the level of the coral reef. This happens by virtue of how the interactions are organized within the reef, which creates interdependencies between individuals' actions and benefits, and solves the problems of commitment, coordination and information exchange.

Future directions

We have argued that biologists working on the functioning and evolution of social behavior stand to gain by adopting the perspective of institutional theory and theoretical tools such as mechanism design; we have also looked at the cases of interspecific mutualisms, antagonistic interactions, group decision-making, breeding systems and signaling games as examples of where and how institutional theory can be of help. In this section, we discuss two other biological questions where the institutional approach promises to be most fruitful.

Cooperative breeding

Cooperative breeding, where individuals other than the parents care for young, is found in many animal species. One of the major theories to explain the evolution of cooperative breeding is called reproductive skew theory (Vehrencamp, 1983), which asks when helpers can breed themselves and how much share of the group's reproductive output they get to receive. The major

models in this field are reproductive transactions, in which one of the individuals (either the dominant, or subordinate) “pays” others with some share of the reproductive output (called staying incentives) in return for helping, and the so-called tug-of-war models, in which all individuals compete with each other for a share of reproductive output. The transactional models can be further subdivided according to whether the dominant individual controls the staying incentives (concession models) or the subordinates (restraint models, Buston et al., 2007). Both of these models are problems in bargaining theory, as has been recently recognized (Cant and Johnstone, 2009), with the individual(s) in control choosing one or the other edge of the bargaining set that maximizes their preferences. In the same vein, (Akçay and Roughgarden, 2007a) modeled reproductive transactions between breeding pairs and proposed using a solution concept, called the core, from cooperative game theory, as a generalization of the reproductive transactions theory. Common to these models is that they assume that the outcome is determined by costless negotiation between the parties (Buston et al., 2007). In contrast, the tug-of-war models have the allocation of reproductive opportunities by scramble competition, which is costly from the group’s point of view, because individuals invest time and resources towards non-productive opportunities rather than producing offspring, causing the reproductive output of the group to shrink.

The tug-of-war and the transaction models represent two extremes of the space of possible institutions. The tug-of-war model represents the lack of any institution, where individuals simply try to grab as much of the proverbial pie as

they can (to maximize their inclusive fitness), whereas the transactions model implies a system of negotiation and communication between the members to determine the outcomes. An institutional approach to cooperative breeding has the advantage of integrating the piecemeal, and sometimes implicit treatments of issues such as how individuals know their outside options, how they communicate these to each other, and what kind of structures enforce the agreements. In addition, mechanism design can generate comparative predictions on possible institutional structures that lie between the tug-of-war model and costless transaction models in terms of the efficiency achieved (and lost).

Within-body mechanism design

Another potentially groundbreaking application of the institutional perspective in biology concerns not animal behavior, but the determination of phenotype by the genetic code. An interesting interpretation of the problem is to treat the phenotype as “agents” of the genes, where genes delegate “decisions” about their fate to the phenotype (the individual organism). In the parlance of principal-agent theory, the individual organism can be considered as an agent with and the gene the principal. Principal-agent theory is concerned with how the principal shapes the preferences of the agent to produce the highest expected benefit for the principal. The gist of the principal-agent problem is that the agent (whose interests are in general not identical to the principal) has private information about either what is the optimal course of action, or what actions it actually undertook (i.e. the principal cannot monitor the agent perfectly). In such

a case, the principal in general has to pay an “information rent” to the agent to induce it to behave in the principal’s best interest (Laffont and Martimont, 2002). In a biological setting, this situation arises because the environment is variable, and the optimal course of action is not known before hand. However, from an evolutionary point-of-view, the “interests” of the agent are not well defined, since phenotypes do not get transmitted to the next generation. Hence, the principal-agent problem between a single genetic principal and the phenotype would be straightforward to analyze, with the gene choosing a preference for the phenotype that maximizes the gene’s expected fitness.

On the other hand, different genes might have different genetic interests, due to differences in their transmission rules (e.g. sex chromosomes, or mitochondrial DNA), a situation called intragenomic conflict (Burt and Trivers, 2006). In such cases, the effect of a given phenotype on the fitness of these different genes might be different, creating a multiple principal situation and making the design on the agent’s preferences a non-trivial problem. Multiple principal problems have been investigated in political science in the context of control of agencies by the legislature, executive branch and outside interest groups (Spiller, 1990). As in the other mechanism design problems, the agency problem created by intragenomic conflict can lead to second-best outcomes relative to what could be achieved in the absence of conflicting interests. Such inefficiencies would also result in selection on transmission rules to minimize inefficiencies, which might be the underlying cause of the evolution of

chromosomes (called a “parliament of genes” by Leigh, 1971), can be understood as such an institution.

Conclusion

The main thesis of this review is that we can use institutional theory to study many phenomena in animal behavior. This approach requires a shift in thinking from the current focus on considering each individual as choosing their strategy alone in a fixed game to considering the organization of a social system that emerges from the interactions between individuals. With this focus, we can interpret the natural history of animal interactions as institutions that encode the timing and manner of how individuals interact with each other. Many institutions in social life organize interactions such that even though individuals rationally follow their self-interest, a measure of efficiency is achieved for the whole system. In the same way, biological institutions can achieve efficiency at an aggregate level without natural selection acting on that level. Thus, an institutional approach to biology promises to be groundbreaking in leading to new questions and answers in our understanding of biological organization.

Acknowledgements

The meeting this paper was based on was held in May 2009 at Stanford University, and was funded by the Woods Institute for the Environment. EA is a postdoctoral fellow at the National Institute for Mathematical and Biological Synthesis (NIMBioS), an Institute 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.

References

- Akçay, E. & Roughgarden, J. (2007a) Extra-pair parentage: a new theory based on transactions in a cooperative game. *Evolutionary Ecology Research*, 9, 1223 -- 1243.
- Akçay, E. & Roughgarden, J. (2007b) Negotiation of Mutualism: Rhizobia and legumes. *Proc. R. Soc. B*, 274, 25 -- 32.
- Akçay, E. & Roughgarden, J. (in preparation) The evolution of games.
- Akçay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. (2009) The evolution of other-regard integrating proximate and ultimate perspectives. *PNAS*.
- Austen-Smith, D. & Banks, J. S. (1996) Information aggregation, rationality, and the Condorcet Jury Theorem. *American Political Science Review*, 90, 34-45.
- Austen-Smith, D. & Feddersen, T. J. (2009) Information aggregation and communication in committees. *Phil. Trans. R. Soc. B*, 364, 763-769.
- Axelrod, R. & Hamilton, W. D. (1981) The evolution of cooperation. *Science*, 211, 1390-1396.
- Baron, D. P. & Ferejohn, J. A. (1989) Bargaining in Legislatures. *The American Political Science Review*, 83, 1181-1206.
- Bowles, Samuel. (2006). *Microeconomics: Behavior, Institutions, and Evolution*. Princeton: Princeton University Press

- Bowles, Samuel, and Herbert Gintis. (2004) "The evolution of strong reciprocity: cooperation in heterogeneous populations." *Theoretical population biology*, 65: 17-28.
- Bshary, R. (2002) Biting cleaner fish use altruism to deceive image-scoring client reef fish. *Proc R Soc Lond B*, 269, 2087-2093.
- Bshary, R. & Grutter, A. S. (2002) Asymmetric cheating opportunities and partner control in a cleaner fish mutualism. *Animal Behaviour*, 547-555.
- Bshary, R. & Grutter, A. S. (2005) Punishment and partner switching cause cooperative behaviour in a cleaning mutualism. *Biology Letters*, 1, 396-399.
- Bshary, R. & Grutter, A. S. (2006) Image scoring and cooperation in a cleaner fish mutualism. *Nature*, 441, 975-978.
- Bshary, R. & Schäffer, D. (2002) Choosy reef fish select cleaner fish that provide high-quality service. *Animal Behaviour*, 63, 557-564.
- Burt, A. & Trivers, R. (2006) *Genes in conflict*, Cambridge, MA, Harvard University Press.
- Buston, P. M., Reeve, H. K., Cant, M. A., Vehrencamp, S. L. & Emlen, S. T. (2007) Reproductive skew and the evolution of group dissolution tactics: a synthesis of concession and restraint models. *Animal Behavior*, 74, 1643-1654.
- Cant, M. A. & Johnstone, R. A. (2009) How Threats Influence the Evolutionary Resolution of Within-Group Conflict. *The American Naturalist*, 173, 759-771.
- Cheney, K. L., Bshary, R. & Grutter, A. S. (2008) Cleaner fish cause predators to reduce aggression toward bystanders at cleaning stations. *Behavioral Ecology*, 19, 1063-1067.
- Clutton-Brock, T. H. (1989) Review lecture: mammalian mating systems. *Proc R Soc Lond B*, 236, 339-372.

- Clutton-Brock, T. H. (1991) *The Evolution of Parental Care*, Princeton, NJ, Princeton University Press.
- Clutton-Brock, T. H. & Parker, G. A. (1995) Punishment in animal societies. *Nature*, 373, 209-216.
- Coase, R. (1937) The Nature of the Firm. *Economica*, 4, 386-405.
- Conradt, L. & List, C. (2009) Introduction. Group decisions in humans and animals: a survey. *Philosophical Transactions of the Royal Society B*, doi:10.1098/rstb.2008.0276.
- Conradt, L. & Roper, T. J. (2005) Consensus decision making in animals. *Trends in Ecology and Evolution*, 20, 449 -- 456.
- Conradt, L. & Roper, T. J. (2009) Conflicts of interest and the evolution of decision sharing. *Phil. Trans. R. Soc. B*, 364, 807-819.
- Dekel, E., Ely, J. C. & Yilankaya, O. (2007) Evolution of preferences. *Rev Econ Stud*, 74, 685-704.
- Enquist, M. (1985) Communication during aggressive interactions with particular reference to variation in choice of behaviour. *Animal Behaviour*, 33, 1152--1161.
- Farrell, J. & Saloner, G. (1988) Coordination through committees and markets. *RAND Journal of Economics*, 19, 235-252.
- Farrell, J. & Simcoe, T. (2009) Choosing the rules for consensus standardization. *mimeo, University of California, Berkeley*.
- Fearon, J. D. (1995) Rationalist explanations for war. *International Organization*, 49, 379-414.
- Feddersen, T. J. & Pesendorfer, W. (1997) Voting Behavior and Information Aggregation in Elections With Private Information. *Econometrica*, 65, 1029-1058.

- Fey, M. & Ramsay, K. W. (2009) Mechanism design goes to war: peaceful outcomes with interdependent and correlated types. *Review of Economic Design*, 3, 233-250.
- Gintis, H. (2003) The hitchhiker's guide to altruism: Gene-culture coevolution, and the internalization of norms. *Journal of Theoretical Biology*, 220, 407-418.
- Gintis, H. (2007) A framework for the unification of behavioral sciences. *Behavioral and Brain Sciences*, 30, 1 - 61.
- Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003) Explaining altruistic behavior in humans. *Evolution and Human Behavior*, 24, 153 -- 172.
- Godfray, H. C. J. (1991) Signaling of need by offspring to parents. *Nature*, 352, 328-330.
- Godfray, H. C. J. & Johnstone, R. A. (2000) Begging and bleating: the evolution of parent-offspring signalling. *Phil. Trans. R. Soc. B*, 355, 1581 -- 1591.
- Grafen, A. (1990) Biological signals as handicaps. *J. Theor. Biol.*, 144, 517 -- 546.
- Greif, A. & Laitin, D. D. (2004) A theory of endogenous institutional change. *American Political Science Review*, 98, 633-652.
- Greif, A., Milgrom, P. & Weingast, B. R. (1994) Coordination, commitment and enforcement: the case of the merchant guild. *The Journal of Political Economy*, 102, 745-776.
- Grossman, S. J. & Hart, O. D. (1986) The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, 94, 691-719.
- Groves, T. (1973) Incentives in teams. *Econometrica*, 41, 617-631.
- Güth, W. (1995) An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *Int J Game Theory*, 24, 323-344.

- Güth, W. & Kliemt, H. (2000) Evolutionarily stable co-operative commitments. *Theory and Decision*, 49, 197-221.
- Kim, Y.-G. (1995) Status signaling games in animal contests. *Journal of Theoretical Biology*, 176, 221-231.
- Klein, B., Crawford, R. G. & Alchian, A. A. (1978) Vertical interaction, appropriable rents and the competitive contracting process. *Journal of Law and Economics*, 21, 297-326.
- Lachmann, M., Számadó, S. & Bergstrom, C. T. (2001) Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences*, 98, 13189-13194.
- Laffont, J.-J. & Martimont, D. (2002) *The theory of incentives: the principal-agent model*, Princeton, Princeton University Press.
- Leigh, E. G. (1971) *Adaptation and diversity: natural history and the mathematics of evolution*, Freeman Cooper.
- List, C. (2004) Democracy in animal groups: a political science perspective. *TREE*, 19, 168-169.
- Maynard Smith, J. & Price, G. R. (1973) The logic of animal conflict. *Nature*, 246, 15-18.
- Mcnamara, J. M., Gasson, C. E. & Houston, A. I. (1999) Incorporating rules for responding into evolutionary games. *Nature*, 401, 368-371.
- Meirowitz, A. (2006) Designing Institutions to Aggregate Preferences and Information. *Quarterly Journal of Political Science*, 1, 373-392.
- Meirowitz, A. & Ramsay, K. W. (2009) The role of international institutions when military capacity is endogenous: an equivalence result. *mimeo, Princeton University*.

- Meirowitz, A. & Sartori, A. (2008) Strategic uncertainty as a cause of war. *Quarterly Journal of Political Science*, 3, 327-352.
- Milgrom, P. R., North, D. C. & Weingast, B. R. (1990) The role of institutions in the revival of trade: the law merchant, private judges and the champagne fairs. *Economics and Politics*, 2, 1-23.
- Myerson, R. B. (1979) Incentive-compatibility and the bargaining problem. *Econometrica*, 47, 61-73.
- Noë, R. & Hammerstein, P. (1995) Biological markets. *Trends in Ecology & Evolution*, 10, 336 -- 339.
- Noë, R., Van Hooff, J. A. R. A. M. & Hammerstein, P. (Eds.) (2001) *Economics in nature: social dilemmas, mate choice and biological markets*, Cambridge, Cambridge University Press.
- North, D. C. (1991) *Institutions, institutional change and economic performance*, Cambridge, Cambridge University Press.
- Ok, E. A. & Vega-Redondo, F. (2001) On the Evolution of Individualistic Preferences: An Incomplete Information Scenario. *Journal of Economic Theory*, 97, 231-254.
- Okasha, S. (2006) *Evolution and the Levels of Selection.*, Oxford, Oxford University Press.
- Ostrom, E. (1991) *Governing the commons: the evolution of institutions for collective action*, Cambridge, Cambridge University Press.
- Pereira, H., Bergman, A. & Roughgarden, J. (2003) Socially stable territories: The negotiation of space by interacting foragers. *The American Naturalist*, 161, 143-152.

- Rodríguez-Gironés, M. A., Cotton, P. A. & Kacelnik, A. (1996) The evolution of begging: signalling and sibling competition. *Proceedings of the National Academy of Sciences*, 93, 14637-146641.
- 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, 965-970.
- Roughgarden, J. & Song, Z. (in preparation) Avian Parental Investment as an Incentive Mechanism.
- Slantchev, B. L. (2003) The power to hurt: costly conflict with completely informed states. *American Political Science Review*, 97, 123-133.
- Spiller, P. T. (1990) Politicians, interest groups, and regulators: A multiple-principals agency theory of regulation, or "Let them be bribed". *Journal of Law and Economics*, 33, 65-101.
- Trivers, R. L. (1971) The evolution of reciprocal altruism. *Quarterly Review of Biology*, 46, 35-57.
- Vehrencamp, S. L. (1983) A model for the evolution of despotic versus egalitarian societies. *Animal Behaviour*, 31, 667 -- 682.
- Williamson, O. E. (1979) Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics*, 22, 233-261.
- Wolff, J. O. & Macdonald, D. W. (2004) Promiscuous females protect their offspring. *Trends in Ecology & Evolution*, 19, 127 -- 134.
- Worden, L. & Levin, S. A. (2007) Evolutionary escape from the prisoner's dilemma. *Journal of Theoretical Biology*, 245, 411 -- 422.

Zahavi, A. (1975) Mate selection - A selection for a handicap. *Journal of Theoretical Biology*, 53, 205-214.