Biological institutions:
The political science of animal cooperation

Erol Akçay[1], Joan Roughgarden[2], James D. Fearon[3], John Ferejohn[3,4], Barry Weingast[3,5]

May 10, 2014

Abstract

The main theme in the evolution of social behavior is how cooperation between organisms can emerge in the face of conflicts of interests. Political science studies cooperation and conflict, and the socio-political structures these produce in the most socially complex animal, humans. In particular, political scientists have long studied how political, economic and social institutions affect human behavior, and how these institutions change over time. We argue that this institutional approach can be applied fruitfully to the evolution of animal behavior. The institutional approach goes beyond the conventional focus on studying the evolution of individual strategies in a given social setting to studying how the social interaction itself is set up. We identify several areas of institutional theory that have immediate applications to biological problems and suggest future avenues for theoretical and empirical research at the interface of social science and biology.

[1] Corresponding author.
Department of Biology, University of Pennsylvani
e-mail: eakcay@sas.upenn.edu
433 S University Ave, Philadelphia, PA, 19146
[2] Department of Biology, Stanford University
[3] Department of Political Science, Stanford University
[4] New York University School of Law
[5] Hoover Institution

Introduction

Human behavior is guided and constrained by social rules, norms and organizations. These rules, norms, and organizational structures, collectively called institutions, emerge in the course of history as a result of past decisions by individuals, groups and societies. Examples include the formation of systems of government, the judicial system, or the regulatory structures of economic activity. But institutions need not to be written in laws and regulations; informal rules and conventions structure human interactions as much as written laws, as anyone relocating to a new country can attest to. A rich theory in political science studies how formal and informal 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 can enable some individual behaviors while constraining others. The result is an interlocking system of social structure.

Animals, of course, lack formal institutions such as legislatures, courts and committees, and the absence of language constrains the complexity of social organization that is possible in animal societies. Nonetheless, animal interactions are also structured by evolved norms and behavioral rules. These norms and behavioral rules determine where and when individuals interact, how long and how close they stay together, what kind of dependency they have on each other, and so on. In short, these norms and rules make up the natural history of an interaction. The main thesis of this paper is that the evolution of natural history can be fruitfully studied from the perspective of institutional theory. We envision a new wave of collaborations between

political scientists and biologists that involves borrowing and adapting tools from each field to study questions in the other. There is a distinguished precedent for such collaboration, as exemplified by the work of Axelrod and Hamilton (Axelrod and Hamilton, 1981) and the recent focus on collective behavior in humans and animals (Conradt and List, 2009).

We call for a shift in perspective on animal behavior. Studies in behavioral ecology traditionally take the context in which animals behave as given, and aim to explain why animals evolve to behave in a particular way based on the fitness consequences of different strategies. In contrast, the institutional perspective we propose focuses on how the rules and norms that structure the interaction evolve.

In keeping with the behavioral ecology tradition that seeks to explain behaviors based on their adaptive value, the first question in understanding the structure of an interaction is to ask what function that structure serves. Political scientists have shown that formal and informal institutions can have different functions: for example, they might allow certain parties (e.g. the ruling elites of a country) to exert coercive control over other individuals. In many cases though, institutions allow *efficiency*, i.e. fully capturing the gains from a social interaction to avoid "leaving money on the table".

To illustrate the concept of efficiency with a biological example, consider two predators sharing a prey killed. An outcome is efficient (from the predators' perspective) if the prey is killed, and all the meat gets eaten. Killing the prey requires some measure of cooperation between the two predators. Even though both predators would prefer the prey being killed and eaten, but they still have conflicting interests between different efficient outcomes – who eats how much – and their conflicting interests may cause them to fail to cooperate. For instance, if the weaker of the two predators expects to be excluded from sharing after the prey is killed, it

might withhold its help during the hunt. 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.

At the heart of our proposal is the hypothesis that the structure of many biological interactions may have evolved so as to sustain mutual benefit. This hypothesis is distinct from group- and kin-selection approaches to the evolution of cooperation, since it does not depend on variation among groups, or relatedness between individuals. Instead, it focuses on the *organization* of the interaction, defined as who interacts with whom, when, and through which proximate mechanisms, and how these features of an interaction may provide direct incentives (i.e., selection through direct fitness benefits) that favor efficient outcomes in an interaction.[6] This in turn leads to the question of how the organization of an interaction evolves as a result of selection acting on the interacting individuals, and if and when such selection can lead to mutual benefit. These questions remain almost completely unexplored in biology and are what we aim to bring into focus with this review.

Commitment, coordination, and private information

The biological literature on the evolution of cooperation has largely focused on games with simple conflicts of interest, such as the prisoners' dilemma or the snowdrift game (Axelrod and Hamilton, 1981, Sachs et al., 2004, Hauert and Doebeli, 2004, Lehmann and Keller, 2006, West et al., 2007). While these studies have elucidated the different ways a direct conflict of interest can be overcome by evolution, other ways cooperation and efficiency can be thwarted have received less attention. In this paper, we review three main strategic problems that can

---

[6] Of course, these direct fitness benefits may act in tandem with either kin- or group-selection, but are logically independent of them.

preclude efficient cooperation, namely, the inability to commit, the need to coordinate and the presence of private information. Each of these problems present distinct challenges, and require distinct solutions.

Institutional theory from political science and economics provides a collection of approaches to address these issues, focusing on the development of social norms and rules that impose constraints on individual behaviors or shape the incentives individuals face. In this review, we illustrate how institutional theory deals with commitment, coordination, and private information problems, and how the institutional perspective can be applied to biology. We then discuss in some detail one of the major tools of institutional theory, called mechanism design, and discuss how an evolutionary mechanism design theory can be developed for biology. Finally, we turn to the role of proximate mechanisms of behavior in structuring social interactions, and the potential role of the institutional perspective in the resolution of the levels-of-selection debate.

Before moving on, a note on terminology is in order. We use the terms "institution", or "institutional perspective" in the biological context to refer to the evolved "rules of the game" of an interaction. Just as the term "strategy" is used in biology refer to alternative behaviors with different fitness consequences without implying conscious deliberation, our use of the term "institution" is not meant to signify any formal organizational structure. Similarly, we use the term "norm" not to refer to any normative prescription (e.g., one *ought* not to steal), but in the accepted game-theoretic sense to refer to patterns of behavior that are self-enforcing (Young, 2008), either directly (e.g., cooperating due to others reciprocating) or through more complex enforcement mechanisms (e.g., costly punishment of defectors).

Commitment

Mediaeval merchants in a coral reef

Mediaeval Europe was a tough place to do business, especially over long distances (Greif, 2006). The rule of law was by no means assured, and property rights and the enforcement of contracts could not be guaranteed by agreements between different polities as they are in today's world. Despite these impediments, mediaeval trade somehow flourished. Greif et al. (1994) tackle this puzzle with a stylized model that asks how a merchant could trust the ruler of a distant city to honor the merchant's property rights and enforce the law. The basic dilemma for the ruler is that even though he benefits from trade (through taxation), he can also rob any merchant that comes to his city. Since any single merchant is unlikely to have great retaliation power, the ruler would do best to rob any given merchant. Expecting this, merchants would be reluctant to visit this city, which would depress the volume of trade, hurting both the merchant and ruler. Thus, the inability of the ruler to commit to enforcing contracts and upholding merchants' rights precludes efficiency.

Greif et al. argue that mediaeval merchants solved this problem by organizing themselves into merchant guilds, which could declare bans on trade with a ruler that cheated one of its members. Thus, if a ruler cheated a member of the guild, he would face retaliation from the whole guild and not a single merchant, so that 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; the commitment problem is solved.

Two points about this model are worth noting: first, the primary function of the merchant guild is not to improve the bargaining position of the merchants to demand, say, a lower tax rate

from the ruler (although subsequently, the guild can do that as well). Rather, the guild enables merchants to coordinate in the face of transgressions, which in turn makes it in the ruler's best interest to honor property rights and agreed upon tax rates. Put another way, the merchant guild replaces the collection of small, individual merchants that do not have retaliation power, and hence act as short-run players, with a single, long-run player that does have retaliation power (Fudenberg et al., 1990). In this way, the guild institution increases the payoff to both the merchants and rulers. Second, while the merchant guild institution solves the ruler's commitment problem, it creates another one, namely that of merchants committing to honor bans imposed by the guild. Thus, the strategic problems faced by the merchants are fundamentally altered through the formation of the guild institution. Greif et al. (1994) argue that complex rules and interrelationships within guilds created incentives for merchants to honor the decisions made by the guild.

A related model by Milgrom et al. (1990) studies the emergence of the Law Merchant to resolve trade disputes in the Mediaeval age. In this model, a given pair of traders interact only once with each other, but have the option of reporting transgressions to a private judge, who keeps records of transgressions that are not remedied. There are efficient equilibria of this game in which all traders consult (by paying a fee) the private judge before trade about whether their prospective partners have outstanding judgments against them, withhold cooperation from those who do, and report any transgression to the private judge after the trade. 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. Note that the private judges do not have any means of enforcement for their judgments. Nonetheless, the institution of the Law Merchant ensures commitment to good behavior by allowing successive trading partners to observe each

other's past behaviors. In repeated games, reputation is one mechanism that allows long-run players to commit to a given course of action (Kreps and Wilson, 1982, Fudenberg and Levine, 1992).

To see how similar ideas can be applied to a biological system, consider 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), which they inspect for ectoparasites and clean. In exchange for their service, cleaners not only consume the ectoparasites, but they can also feed on the healthy mucus and scales of their clients. Since consuming healthy tissue hurts the clients, but is preferred by the cleaners, there is a conflict of interest between the partners in the mutualism (Bshary and Grutter, 2002).

There is substantial research on how clients keep cleaners from cheating and feeding on healthy tissue. Different client species seem to have different options available to them: some are predatory and can in principle eat a cheater cleaner fish while others can simply choose to associate with more cooperative cleaners (Bshary and Schäffer, 2002). Yet another class of clients resorts to a punishment strategy, chasing the cleaner after being cheated (Bshary and Grutter, 2002). Clients also observe other clients interacting with their prospective cleaner so that there is reputation effects involved (Bshary, 2002). The combination of retaliation strategies and reputation effects makes it in the cleaner's best interest not to cheat clients. Remarkably, cleaners put in novel situations in laboratory experiments are able to learn new retaliation rules, and adjust their behavior in order to optimize their gains (Bshary and Grutter, 2005, Bshary and Grutter, 2006).

These rules represent a collection of simple informal institutions that together give rise to

a biological market (Noë and Hammerstein, 1995, Noë et al., 2001, Bowles and Hammerstein, 2003). The aggregate effect of these rules is that over time the more cooperative cleaners will be preferred by clients and thus receive more "business", which can maintain cooperation in this system. Hence, the market institution allows mutually beneficial exchanges in the absence of a system of laws and courts.

However, as economists have known for a while, markets can be inefficient in the absence of institutions that provide for enforcement of contracts and the availability of reliable information about the quality of service and prices (North, 1991). Similarly, for the cleaner fish market to work efficiently, clients should be able to reliably assess the "price" (e.g. healthy tissue eaten per parasite removed) each cleaner charges and cleaners need to be able to commit to their "advertised" prices. These two features cannot be automatically assumed, as illustrated by the presence of cleaners who signal clients their cooperativeness, but then go on to feed on healthy tissue (Bshary, 2002). Market choices can prevent or limit such transgressions under some conditions. However, when cleaners are saturated, the cost of losing any single client will be small, and hence cleaners will not be able to commit not cheating any particular client, similar to the rulers of mediaeval cities (Greif et al., 1994). This will depress visitation rates by clients; the market will not function with full efficiently. Furthermore, a client fish will be limited to information that it can directly gather by interacting with a cleaner or observing other clients' interactions. Obtaining information in this way is likely to be costly and easily manipulated, which also will limit the efficiency of a market.

An institution that keeps track of the long-term performance of a cleaner can solve these problems. In particular, one might conjecture that some small territorial clients, who are long-term associated with a single cleaning station, can serve as repositories of the cleaner's past

performance. 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 (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, which makes it in the cleaner's best interest to not cheat and solves the commitment problem. This equilibrium also changes the incentives between the territorial clients and cleaners, and would prompt cleaners to be cooperative towards their territorial clients. Empirically, such an arrangement would predict that the interactions between territorial and roaming clients should play an important role in maintaining cooperation in the cleaner fish system.

Note again that the argument above focuses on the structure of the game. It considers how the incentives that the cleaner fish and its clients face change when a third party is introduced, and whether this can lead to a higher level of efficiency at the level of the interaction. In other words, it is about the evolution of games organisms play, rather than the strategies within a given game. Such analyses are the hallmark of institutional theory in political science, but are rarely done in biology (but see Worden and Levin, 2007, Akçay and Roughgarden, 2011)

The theory of the firm and breeding systems

An important area of economics where commitment problems play a central role is the theory of the firm. A firm in economics is defined as 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 ownership of production assets and decision-making in some agents (the employer) and remunerate others (the employees) in return. One of the main questions in the theory of the firm is why firms exist at all, i.e. why do agents organize themselves into long-term relationships as

opposed to achieving production by on-the-spot transactions in the marketplace? This question, first posed by Ronald Coase in his influential essay (Coase, 1937), has stimulated a large body of research in law, economics, and political science.

Coase 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 the terms of agreement. A more subtle type of cost occurs when production assets are specific to each other (e.g. a supplier needs to build a plant next to a manufacturer's factory). In such cases, recourse through partner choice in the open market may become infeasible. In particular, situations that are not covered in the original agreement between parties (e.g. unexpected movements in prices) would 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 of a strong and weak predator cooperating to hunt prey. The inability of parties to commit to not taking advantage of each other can then lead to underinvestment into the production assets 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. (1978) argue that efficient transactions can be achieved in such cases when one party owns both assets, instead of relying on repeated market transactions. Likewise, Grossman and Hart (1986) show that if investments are relationship-specific and contracts cannot account for all possible contingencies, the problem of underinvestment can be solved by assigning control of both assets to one party (the firm), which removes the incentive to appropriate sunk costs after the investments are made. In a related vein, 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 (e.g. partners, or employees and employers), rather than particular actions they need to take in each possible state of the world.

How can these ideas be applied in biology? Consider the striking contrast between the breeding systems of birds, where the overwhelming majority of species exhibit social monogamy (i.e. a male and a female raising offspring together), versus those of mammals, where most species are polygamous. Previous thinking on this pattern mostly departs from the assumption that females prefer monogamy, and males polygyny (Clutton-Brock, 1991), and asks when males are constrained to be monogamous, either through female choice, or due to females' distribution in space and time (Orians, 1969, Emlen and Oring, 1977, Clutton-Brock, 1989).

Viewing offspring rearing as analogous to the production of goods by a firm leads to a new perspective that predicts not only the occurrence of monogamy versus polygamy, but also the nature of the relationships between the mates. 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. Roughgarden (2009) proposed that polygyny may evolve because males compensate for being excluded from making investments directly into the offspring.

An alternative hypothesis can be constructed by observing that most mammals, females are mobile during gestation and offspring are either mobile shortly after birth, or can be carried around. Hence, mammal females can receive help from different males, for example, by moving between territories or associating with different males, without necessarily losing their offspring (especially if they have copulated with multiple males, Wolff and Macdonald, 2004). A female bird with altricial young, however, cannot move her eggs or nestlings to another nest; likewise,

the male cannot share the nest and the territory he has invested in with another female while another brood is in it. Hence, investments by the male and female into breeding are relationship-specific in most bird species, whereas they are far less so in mammals.

Under these crudely generalized conditions, the theory of the firm predicts that interactions between mates in birds should be governed by "relational contracts", i.e. longer-term commitments. The pair-bond of birds would be analogous to a relational contract, if it proximately alters individuals' behavioral tendencies to act outside their short-term selfish interest (Akçay et al., 2009) and according to the needs of the brood (Akçay and Roughgarden, 2009). Of course, for such proximate commitment mechanisms to evolve, parties need to be able to reliably ascertain each other's commitment before the breeding attempt. Extensive courtship behaviors in birds can be hypothesized to serve this function (in addition to determining the likely quality of parental care supplied by the male; Hoelzer, 1989). In contrast, 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. This argument suggests that transitions between monogamy and polygamy is possible even if the resource requirements and distributions of females stays the same (Emlen and Oring, 1977, Clutton-Brock, 1989). The theory-of-firm hypothesis thus holds that monogamy and social pair bond will evolve not because of the inability of males to monopolize females, but because of the inability of the males and females to switch partners without losing their past investments. This prediction can be tested by taking advantage of the variation in offspring mobility in different mammal and bird species: it predicts that, regardless of home-range sizes, species with more mobile offspring will be more likely to have social polygamy.

Conversely, species where offspring are less mobile will be more likely to exhibit social monogamy.

Coordination

How to agree on industrial standards

Many social and economic situations are characterized by the need to coordinate to achieve mutual benefit. Coordination problems arise in settings that range from competition between different industrial standards to the movement decisions of an elephant herd. Even in cases where the interests of the parties are largely concordant, coordination failure can preclude efficiency.

In an influential paper, Farrell and Saloner (1988) compare two different institutions through which two players can coordinate their actions. The first is a committee where players negotiate with each other, while the second one is a market mechanism where both players come forward with their own actions and hope that the other follows suit. The committee institution ensures coordination, but imposes negotiation costs (in particular, delays in agreement), while the market minimizes delay costs but creates the possibility of mis-coordination if players commit to different actions at the same time. Farrell and Saloner show that a hybrid institution that prescribes negotiations 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 (2009) study when, in the context of deciding industrial standard, the use of costly tactics to persuade the other party to give up (i.e., playing a war-of-attrition game) will be optimal 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 most efficient 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. Interestingly, both players prefer such an institution to the unchecked war-of-attrition before they learn their own proposals' quality, but they might prefer continuing with the unchecked war-of-attrition game after they have learned this information. This means that a mechanism needs to be in place to commit both sides to accept the randomly chosen outcome in case of persistent disagreement.

A biological coordination problem can be found in collective decision-making with multiple alternatives, an area that has seen a rapid growth in recent years (Conradt and Roper, 2005, Conradt and List, 2009). The results of 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 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 (2009) 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 randomize the decision. For example, the individual in best condition in the group (e.g. the most satiated and hydrated) can arbitrarily make the decision, since such an individual would represent the closest approximation to a neutral party in the group. In general, collective decision in humans and animals making also presents problems of private information, which are taken up below in the subsection "Voting and information aggregation".

Private information

Perhaps the most fundamental reason precluding full efficiency in an interaction is that parties have private information about important variables affecting their payoffs. Informed parties will often have incentives not to reveal their information truthfully, even when their interests are completely concordant. Therefore, special incentive schemes are needed in order to get individuals to reveal their information. We will illustrate how this is achieved using the examples of international conflict and voting.

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 simple observation that wars are inefficient. Wars destroy valued resources and often pose significant risks to the political leaders who start them, so given any outcome of the war, both sides 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 the private information each state has about its own attributes, such as its military capability or the value it places upon the disputed territory, will preclude finding a mutually acceptable peaceful settlement (Fearon, 1995). In particular, when two states are engaged in pre-war negotiations, both might have incentives to withhold or misrepresent their private information, either because they expect 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 strength and the value it places upon the object of contention, no feasible negotiated outcome may exist that looks preferable to war for both states.

War may then follow as a means of credibly revealing (or bluffing about) one's private information. However, under some circumstances, practices in international "militarized disputes" that are short of a full-blown war – such as mobilizing troops and issuing public statements that have domestic consequences – may also enable credible revelation through "costly signaling."

Another way that a full-blown war might be prevented is through international institutions that give states incentives to truthfully reveal their private information. Recent literature on the political science of war asks what is the best international institutions can do in reducing the probability of war (e.g. Fey and Ramsay, 2009, Meirowitz and Sartori, 2008). One important result from this literature is that when states' private information is correlated with each other (e.g., when the information is about military capabilities) then regardless of the details of the negotiation process between the states there is a positive risk of fighting, provided that the total costs war are not too high. This is similar to the foundational result of Myerson and Satterthwaite (Myerson and Satterthwaite, 1983) in the setting of bilateral trade between buyers and sellers with private information, which states that there exist no trading scheme that guarantees efficient trades to always happen as long as there is some probability of inefficient trades. Both of these results are obtained using a powerful theorem from mechanism design theory, called the revelation principle (Myerson, 1979, see also the section on mechanism design below), and put limits on how efficient the outcome of an interaction can be made given that there is private information.

Biologists have long used the metaphor of war and peace for agonistic behavior between animals (Maynard Smith and Price, 1973), and there are some close parallels between previous theory in animal behavior and political science. Maynard Smith and Parker (1976) were among

the first to recognize that private information can lead to escalated conflicts. Subsequent studies (Enquist, 1985) (Maynard Smith and Harper, 1988) showed that when two animals have private information about their valuation of the resource or their fighting ability, evolutionarily stable, costless signals can exist and allow animals to avoid fighting some of the time. In this case, the honesty of these signals is maintained not through intrinsic costs, but because individuals who pretend to be stronger than they are face fights with stronger individuals. Individuals who pretend to be weaker, on the other hand, experience opportunity costs due to the contests they could win without fighting, but have to fight out because of their signals.

The focus in biology on how animals avoid fighting costs parallels the political science literature on crisis bargaining and international institutions to prevent war, but with some important differences. First, biological models tend to focus on simple games such as the Hawk-Dove game, and usually do not allow negotiated partitions of the contested resource through pre-fight interactions. In many interactions, however, the contested resource can be divided, such as territorial interactions (Stamps and Krishnan, 2001, Pereira et al., 2003) or bargaining over resource exchange (Akçay and Roughgarden, 2007). Bargaining models commonly used in the theory of social and political institutions can therefore be used to extend existing biological theory to cases where a near-continuous division of the resource is plausible. Furthermore, a mechanism design approach can help to extend these results to generate "game-free" results about the expected frequency and level of conflict (Fey and Ramsay, 2009).

Voting and information aggregation

How groups of individuals make decisions is one of the central questions in both political science and behavioral ecology and has recently seen fruitful collaborations between the two disciplines (List, 2004, Conradt and List, 2009). Voting theory is concerned with how and

whether efficient outcomes can be achieved when multiple individuals have to make a collective decision. The social science literature usually takes the point of view of a "social designer", whose goal is to satisfy group-level criteria, such as efficiency, for the aggregate decisions. In contrast, biologists usually focus on individually optimal strategies in a given game (Conradt and Roper, 2009).

An interesting problem in voting theory arises when individuals have private information about their own preferences or the environment. Suppose for the moment that individuals share a common interest but have independent estimates of what action best furthers the common interest. If all individuals have the same and better-than-even-chance of being accurate, then consulting more individuals and taking the average of all estimates, on average, improves the accuracy of the decision. This result is known as the Condorcet Jury Theorem. Problems arise, however, when group members decide strategically about whether or not they will reveal their information truthfully, a behavior called strategic voting. Surprisingly, in such cases it may pay for individuals to withhold or misrepresent their information, even with complete concordance of interest (Austen-Smith and Banks, 1996, Austen-Smith and Feddersen, 2009). The reason is that when deciding how to vote, a rational individual should consider only the case where he or she would be casting the deciding vote; in all other cases it doesn't matter what he or she does. But the event that the focal individual's vote is pivotal implies that the other individuals are voting in a particular way (e.g. under unanimity rule, all others must be voting "Yea" for the focal individual to be pivotal). Hence, the focal individual should condition her vote upon that state of the world, which might make voting against one's private signal optimal. Therefore, strategic voting is frequently not informative about the private information of agents. Generally speaking, 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 more significant problem in smaller groups, since each individual has a higher probability of being pivotal. Conversely, in large groups the Condorcet Jury Theorem approximately survives strategic voting (Feddersen and Pesendorfer, 1997).

Not surprisingly, the problem of strategic voting and misrepresentation is aggravated when there are real conflicts of interests within the group. However, Meirowitz (2006) shows that outside transfers to individuals as a function of their revealed information can create such incentives, and that the magnitude of the required transfers becomes smaller as group size gets larger (due to, again, 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, as we discuss below.

Mechanism design, auctions and costly signaling

We now turn to mechanism design theory and the related theory of auctions. Conventional game theory starts by specifying a game structure, and predicts outcomes supposing self-interested agents with some level of computational capacity and access to specified public and private information. Mechanism design inverts this approach: it specifies the information structure and a set of possible games, and finds the game that produces the outcome with a desired property, such as achieving efficiency. Auctions are particular instances of this problem where the question is how to sell an object to a buyer in a way that maximizes the profit to the seller.

The term "mechanism" in game theory is used to describe the "rules of the game"; in other words, mechanisms are mathematical representations of an institution. In contrast, biologists use the term "mechanism" to refer to the processes that bring about a behavior, i.e. the

"proximate cause" of behavior, such as the firing of specific neural circuits in the brain or changes in hormone levels. To avoid confusion, we use the term "mechanism" by itself in the game theoretic sense, and refer to biological mechanisms with the phrase "proximate mechanism". But the two senses of the term mechanism are intimately connected: game-theoretic mechanisms represent the structure of the interaction generated by proximate mechanisms of behavior.

Mechanism design theory primarily deals with situations where agents have private information and face incentives to not reveal this information accurately, as in the cases of international conflict and voting discussed above. Some mechanism design models take up the perspective of a player who has the power to alter the game structure to better serve its own interest, for example a parent company dealing with subsidiaries in a conglomerate (Groves, 1973) or an auctioneer who tries to set up an auction that provides the most revenue from the sale of an item (Vickrey, 1961, Klemperer, 1999). Other mechanism design problems presume a "social designer" that is interested in improving aggregate welfare, for example, an international organization trying to prevent wars (Fey and Ramsay, 2009) or a government setting up a tax schedule to provide public goods (Clarke, 1971). Regardless of the objective of the designer, the focal issue in mechanism design is to set up a game that makes truthful revelation of information individually optimal, i.e. incentive compatible. Fundamentally, incentive compatibility is means aligning the interests of a self-interested agent with whatever the objective of the mechanism is. This frequently requires payments to be made to, or deducted from, the individuals as a function of the messages they reveal.

Incentive compatible mechanisms are crucial to achieve cooperation in diverse situations such as bargaining problems (Myerson and Satterthwaite, 1983) or producing public goods

(Groves and Ledyard, 1977), where the mutually beneficial policy will depend on the private information of the individual agents. One general result from mechanism design is that ensuring incentive compatibility frequently leads to ineluctable costs when there are conflicting interests between agents (Myerson and Satterthwaite, 1983). The difference between the overall payoff that could have been achieved if there was no private information and that achieved under an incentive compatible mechanism is sometimes termed "agency loss".

In biology, private information problems have been investigated in three main contexts: agonistic behavior (Maynard Smith and Parker, 1976, Enquist, 1985; see also "International conflict and how to prevent it" above), signaling of mate quality (Zahavi, 1975, Grafen, 1990), and signaling of need between parents and offspring or other relatives (Godfray, 1991, Johnstone and Grafen, 1992, Nöldeke and Samuelson, 1999, Godfray and Johnstone, 2000). The method of analysis in most of these models is to assume that there is a costly behavior or trait -- the signal -- that the informed party (e.g., males who know their quality, or offspring who know their need) can invest in. The honest (or separating) signaling equilibrium has to satisfy two conditions: (1) at the optimal signal level of the signaler given the cost of signaling and the benefit resulting from the uninformed party's reaction to the signal, is different for different values of the private information, (2) assuming that the signaler has signaled truthfully, the receiver reacts to the signal to maximize its own payoff. Conditions (1) and (2) imply that incentive compatibility is satisfied for the signalers and the receiver has gained full information. In all these models, costs of signaling are wasted, reflecting agency loss. In fact, the agency loss in parent-offspring signaling can be so high as to render one or both parties worse off relative to no signaling (Rodríguez-Gironés et al., 1996, Johnstone, 1999, Godfray and Johnstone, 2000).

What can mechanism design theory do for biology? We see potential contributions of the mechanism design approach in terms of methodology and increasing the scope of strategic issues considered in models.

In terms of methodology, mechanism design provides powerful and well-developed techniques to characterize and solve incentive compatibility problems. Chief among these techniques is the revelation principle (Myerson, 1979); which states that any (Bayesian Nash) equilibrium of any game with incomplete information can be represented as an equilibrium in a special class of game called direct mechanisms. A direct mechanism is a game where all individuals reveal their information truthfully to a (imagined) central arbiter, who then decides what outcome of the game is implemented and which payoff each player will receive based on the revealed information. To implement the payoffs, the direct mechanism may make payments to the players or impose costs. The revelation principle allows one to answer many questions in mechanism design, such as determining the best outcome that can be achieved under incentive compatibility constraints, by focusing on direct mechanisms only which allows generating very general results for a large class of games with imperfect information.

In the first application of mechanism design to biology, Akçay et al. (2012) consider the phenomenon of reproductive skew, which is the question of how the reproductive output of a cooperatively breeding group is to be distributed among the group members (Vehrencamp, 1983). This question is of fundamental importance in the evolution of sociality, since it determines when selection favors forming groups or not. The classical theoretical framework to understand reproductive skew is called "reproductive transactions theory" (Vehrencamp, 1983), which conceives of fitness transactions within a group so as to make all members at least as well off (in terms of inclusive fitness) when joining the group relative to each individual's outside

option (i.e., being solitary). The simplest model of reproductive transactions has two individuals that are related to each other, and that have specified outside options and a total group output (also specified) and finds the range of divisions of the group output that makes both individuals better off if they give up their outside options and join the group (Johnstone, 2000). In general, this region is larger when individuals more highly related to each other. Reproductive transactions theory suffers from two limitations: one is that it makes strong perfect information assumptions: individuals (or natural selection) can condition decisions on all of outside options, group output, relatedness, etc.. In particular, thanks to perfect information, the classical theory predicts that groups will form whenever they are efficient (i.e., the group output is greater than the sum of the outside options). The second, and related limitation is that the predictions of reproductive transactions theory only give a range of outcomes compatible with evolutionary stability, which creates a sensitive dependence on model details as to which outcome from this range will be selected. In this way, classical reproductive transactions theory is quite similar to cooperative game theory, and the range it predicts is akin to the core (Aumann, 1989). Akçay et al. (2012) modify the reproductive transactions framework by making the outside options of individuals private information. With this modification, the 2-person reproductive transactions game has the exact same structure as the canonical bilateral trade model with private valuations (Myerson and Satterthwaite, 1983), except that individuals are selected to maximize their inclusive fitness (or equivalently, have other-regarding preferences). Akçay et al. show that the seminal inefficiency result of Myerson and Satterthwaite extends to this case: unless relatedness is above a certain threshold, some pairings that are efficient will not happen at the evolutionary equilibrium under any mechanism. This suggests that natural selection might act on the mechanisms to minimize the inefficiency, since this benefits all individuals.

Another foundational result obtained from the revelation principle is the revenue equivalence theorem, which states that when buyers in an auction have independent valuations about an object, many commonly used auction mechanisms (such as the English and Dutch auctions) yield the same expected revenue to the auctioneer (Vickrey, 1961, Myerson, 1981, Riley and Samuelson, 1981). Results such as revenue equivalence can be applied to biology as exemplified by the case of multiple offspring signaling their need to the parent. Johnstone (1999) analyzed what amounts to an all-pay auction where multiple offspring "bid" for a food item by engaging in costly begging, and finds the signal costs that make truthful begging optimal. The revenue equivalence result implies that the expected costs will be independent from the details of the signaling mechanism as long as the item, if awarded, is given to the highest bidder. Hence, one can analyze computationally simpler mechanisms to determine the expected costs. More importantly, one can use an analysis similar to Myerson's (Myerson, 1981) to determine the optimal auction rule that maximizes the profit to the parent (i.e., the auctioneer). The difference with the auction models in economics is that an auctioneer wants the bidders to pay a higher cost (as this is his income), while the costs in parent-offspring signaling are deducted from the parent's fitness.

Apart from the methodological contributions, mechanism design and auction theory also suggest a number of new strategic problems that have biological significance. One such problem that arises in auctions is when the bidders do not have entirely independent valuations of the items offered. This so-called "common values problem" arises, for example, in auctions for oil drilling leases, where the value of the lease to all players depends on how much oil there really is beneath the soil, but different players might have different estimates for how much oil there is, and may have different costs (due to technological or logistic reasons) for extracting and

processing it. A similar problem arises in biological signaling where the signalers do not know the objective quality of the item offered; for example, offspring may not know how nutritious a given food item is, or different males courting a female might have different estimates of the breeding ability of a female. In these cases, the value of "winning" the auction to the signalers will not be independent. In general, revenue equivalence breaks down for common value auctions (Milgrom and Weber, 1982); that is, different auction mechanisms produce different expected proceeds to the auctioneers. Furthermore, a common phenomenon associated with common values auctions has been termed the "winner's curse", where the fact that a bidder has won an auction on average means that he has overestimated the real value of the item. Hence, optimal bidding in common values auctions usually involves individuals "shading" their bids, i.e. bid lower than their actual valuations, which reduces the revenue to the auctioneer. Therefore, the auctioneer can sometimes improve its expected payoff to reveal whatever information it has truthfully (Milgrom and Weber, 1982). However, revenue in economic models corresponds to wasted signal costs in biological signaling; which suggests, for example, that parents in a costly begging setup might in fact do better to hide the real value of the food item from the offspring, so as to decrease the incentives to engaging in excessively costly begging. This, and other strategic complexities arising from common values remain to be explored.

Finally, an understudied issue in biological signaling is the role of behavioral dynamics in determining the outcome of signaling interactions. Existing biological models of signaling for the most part consider both the signaler and receiver behavior as evolving at the evolutionary time-scale. In reality many of these interactions involve the parties reacting to each other in the behavioral time-scale and are able to "learn" what different signals mean or what reactions they elicit (Rodríguez-Gironés et al., 2002, Grodzinski et al., 2007). Two recent studies, by

Roughgarden and Song (2014), and Akçay (2012) model parent-offspring communication in the behavioral time-scale. Roughgarden and Song consider a supply-demand dynamics between the parent and the offspring, with the parent adjusting the "price" of food (i.e., the amount of begging the parent requires for unit food delivery), and the offspring adjusting its demand in response to the price it sees. Roughgarden and Song consider the "market-clearing" equilibrium of this dynamic where the offspring demand exactly the amount of food the parents are willing to provide, and show that this equilibrium is separating. Akçay, on the other hand, models the communication process as a dynamic auction process, similar to the canonical model of mechanism design, the Vickrey-Clarke-Groves auction (Vickrey, 1961, Clarke, 1971, Groves, 1973). Akçay assumes that the parent has a genetically determined decision rule specifying at which rate food is yielded in response to offspring begging. The offspring "learn" to beg optimally given the parent's decision rule, leading to a separating equilibrium. Akçay then considers the evolution of the parent's decision rule, and shows that under some parameter values, the parent can evolve to let the offspring demand food without any cost. The behavioral equilibrium induced by this decision rule is still separating, since offspring have a partial interest in the parent's survival, and therefore lower need offspring demand food at lower levels. In the language of mechanism design theory, the parent's decision rule corresponds to a mechanism, and the evolution of the decision rule to finding the optimal mechanism.

Proximate causes, institutions and the levels of selection

In the previous sections, we surveyed some important areas of institutional theory and suggested parallels and applications of some of these ideas in biology. We used the term "institution" for a theoretical construct representing regularities and constraints in social and

political interactions. At the end of the day, though, such regularities and constraints are the result of real-world mechanisms. In political and economic interactions, these mechanisms can be very abstract, and enshrined in laws, treaties and contracts that provide for enforcement of a seemingly endless array of possible rules. In contrast, non-human animals lack language and other tools for implementing formal institutions, and hence many mechanisms of external enforcement are not available to animals. Instead, the rules and conventions that structure animal social interactions come about through how individuals make decisions based and react to each other; in other words, the proximate mechanisms of behavior. These proximate mechanisms determine both the incentives individuals face in social interactions, as well as how they react to these incentives.

The closest thing in standard economics to a proximate model of behavior is the rational actor model (RAM). The RAM specifies that agents have beliefs and expectations over the state of the world and their partner's types, actions, etc., and can carry out the necessary calculations to find actions that maximize their utility (however defined) given these beliefs and act accordingly. The RAM underlies widely used solution concepts such as the Nash or Bayesian Nash equilibria. Biologists, on the other hand, have traditionally tended to work with very simple models of proximate causation, where the behavior of an individual is either under direct control of a genetic locus, or has very simple conditionality. Increasingly however, both fields have been moving away from their traditional models. The RAM has been extended to account for constraints on computational capabilities and cultural influences on preferences (Gintis, 2007), while biologists have been developing game theoretic models with explicit mechanisms of proximate causation (McNamara et al., 1999, Taylor and Day, 2004, Roughgarden et al., 2006, Akçay et al., 2009, Roughgarden, 2009).

Of these recent models in biology, the one by Akçay et al (Akçay et al., 2009) provides many immediate linkages to economic theory and a long line of thinking in behavioral ecology that views animals as goal-oriented agents (McFarland and Houston, 1981). In this model, individuals have a genetically encoded objective function (representing the reward sensation of the agents) and act myopically to maximize these objective functions. These objective functions assume a role akin to the utility function in a boundedly rational actor model. The behavioral dynamics of Akçay et al. (2009) result in a pure strategy Nash equilibrium of a game defined by the objective functions, which can take into account things other than the material payoffs of individuals. The equilibrium actions of the individuals determine individuals' fitness, and result in natural selection acting on the objective functions. Using this framework, Akçay et al. show that other-regarding objective functions can evolve even in the absence of kin- or group-selection, and under complete conflict of interest between the players. This result provides partial support for the conjecture by Roughgarden et al. (Roughgarden et al., 2006, Roughgarden, 2009) that social behaviors might be driven by non-competitive dynamics such as the "team-play dynamics", which leads to the Nash Bargaining Solution (Nash, 1950) as a behavioral equilibrium. The model of Akçay et al. is also related to "indirect evolution" models in economics (e.g., Güth, 1995, Weibull and Salomonsson, 2006, Dekel et al., 2007), which consider the evolution of individuals' preferences in social interactions.

Non-selfish objectives, or preferences, can be interpreted as internal commitments to cooperate against the actor's short-term material interest (Güth and Kliemt, 2000). This mechanism of commitment contrasts with the model of the Merchant Guild (Greif et al., 1994), which relies on pure self-interest of the actors. 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).

The significance of such internal commitments is that they change the incentives that the social partners of the focal individual face. In Akçay et al's model (2009), other-regard evolves because individuals that face an other-regarding individual become better off helping the focal individual, which in turn increases the fitness of the focal individual. Similarly, in the costly signaling models of parent-offspring communication (Godfray, 1991, Johnstone, 1999, Akçay, in review), the proximate mechanisms that underlies the parent's allocation rule and the offspring's decisions to signal determine how much the offspring beg for food and what costs they incur. As such, we expect there to be direct selection on these proximate mechanism of decision-making, through which the game theoretic mechanism ensuring truthful communication can evolve. More recently, Akçay and Roughgarden (2011) consider how traits that provide explicit incentives to others to cooperate can evolve, and find that evolution can lead to polymorphisms in the incentives provided, which sustains much cooperation in the population, but also generates a complex pattern of behavioral diversity.

Another example of proximate mechanisms changing incentives is the idea that punishment of cheaters that can maintain cooperation by making cheating materially unprofitable. Yet punishing can be materially costly and individually suboptimal; hence Gintis and colleagues (Gintis et al., 2003, Gintis, 2003) argue that a commitment to punishing is achieved through the evolution of other-regarding preferences. Thus, simple informal institutions that provide material incentives and disincentives can be implemented through proximate mechanisms of behavior determining individual decision-making.

Thus, an increasing number of studies investigate the idea that social interactions are structured by evolution changing incentives that parties face in behavioral time-scales, but much remains unexplored. We believe that this issue represents one of the major areas of development for evolutionary biology. The institutional perspective, coupled with models that integrate proximate mechanisms in evolutionary models hold great promise for tackling these questions.

Institutions and levels of selection

Finally, the institutional perspective we advocate in this paper puts a new twist to an ancient, but still ongoing, debate in evolutionary biology about whether some social adaptations can sometimes be understood as serving the "good of the group" (Okasha, 2006, Gardner and Grafen, 2009). For example, consider a coral reef with many interacting species, including the cleaner wrasse and its clients. Neither kin selection nor group selection can act on a coral reef ecosystem as a whole, 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, it is possible for natural selection to shape the organization of the interactions in a coral reef though acting on individual species' traits, and align selection at the individual level with the objective of maximizing the mutual benefit at an aggregate level. To be sure, this is by no means an inevitable outcome of natural selection, yet at the same time, it represents an understudied possibility in evolutionary theory. The institutional perspective provides the theoretical framework to investigate when and how efficiency at the aggregate level can be attained in the face of conflicting interest, and problems of commitment, coordination and private information.

Recently, Bowles et al. (2003) modeled how group-level institutions in humans (such as taxes to share benefits and intra-group assortment mechanisms) can evolve through group selection (either cultural or biological). In their model, institutions evolve because they facilitate

the increase of group-beneficial behaviors within a group. Thus, groups that possess institutions can outcompete other groups in direct conflicts, and replace them, leading to increases in group-level efficiency. This argument is quite plausible for early human societies for which it was proposed, but perhaps not so for the majority of animal interactions that do not take place in well-defined groups. However, natural selection can still align the interests of the individual with that of the group in the absence of clear group structure. In particular, individuals' behavioral responses to each other can provide immediate incentives for cooperation. Recently, Akçay and Van Cleve (in review) have provided the general conditions for when behavioral responses in structure populations can evolve to maintain group-optimal behaviors. In particular, when individuals behavioral responses completely match each other's actions, the interests of individuals align completely with each other, and with the aggregate group payoff. Akçay and Van Cleve show that such completely matching behavioral responses can emerge through the evolution of prosocial preferences in public goods games, without the need for group selection. Once these prosocial preferences evolve, they render group-optimal behavioral outcomes evolutionarily stable because they motivate individuals to match each other's contributions. This provides the simplest possible form of institution to maintain efficient outcomes, consisting simply of the evolved norm of prosocial preferences, and relying on individual's immediate responses to each other. However, even in situations where immediate responses are not possible (due to separation of acts in space or time), more complex elements of social structure (such as the equilibrium of the law merchant described above) can still arise through selection acting on individuals, and have the effect of sustaining group-level efficiency. The application of the law merchant model to cleaner fish that is discussed above is an example of this: each party acts in their own self-interest given what the others do, yet the arrangement of interactions solves the

informational and commitment problems and maintain highest mutual benefit. In short, biological "institutions" might allow group-level optimization without group-level selection.

Conclusion

The main thesis of this review is that the theory of political and economical institutions can be profitably utilized to study many phenomena in animal behavior. We did not attempt to give a complete overview of institutional theory, which deals with diverse questions that range from the structure and functioning of the legislative process (Baron and Ferejohn, 1989), to economic development (North, 1991) and management of common resource pools (Ostrom, 1991). Rather, we highlighted some major areas to illustrate the approaches taken in the field and how they could be applied to biology. What distinguishes institutional theory from the rest of game theoretic inquiry on economics and social science is not a unique central theorem, but rather an approach that seeks to understand systems of interaction as designed or evolved solutions to underlying problems of exchange, production, or allocation. To use such an approach in biology, we need to shift our perspective from current focus on considering each individual as choosing a strategy alone in a pre-determined game, to considering the organization of a social system that emerges from the interactions between individuals. From this perspective, the natural history of animal interactions can be seen as similar to institutions, encoding the timing and manner of how individuals interact with each other, and potentially functioning to facilitate the evolution of mutual benefit through selection acting on individuals.

References

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

Akçay, E. & Roughgarden, J. 2007. Negotiation of Mutualism: Rhizobia and legumes. *Proceedings of the Royal Society B,* 274**,** 25 -- 32.

Akçay, E. & Roughgarden, J. 2009. The perfect family: biparental care in animals. *PLoS ONE,* 4**,** e7345.

Akçay, E. & Roughgarden, J. 2011. The evolution of payoff matrices: providing incentives to cooperate. *Proceedings of the Royal Society B,* 278**,** 2198-2206.

Akçay, E. & Van Cleve, J. 2012. Behavioral responses in structured groups pave the way to group adaptation. *The American Naturalist*; 179:257-269

Akçay, E., Meirowitz, A., Ramsay, K. & Levin, S. 2012. Evolution of cooperation and skew under imperfect information. *Proceedings of the National Academy of Sciences*, 109, 14936-14941

Akçay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. 2009. The evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences,* 106**,** 19061-19066.

Aumann, R.J. 1989. *Lectures on Game Theory*. Boulder, CO: Westview Press.

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. *Philosophical Transactions of the Royal Society 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, S., Choi, J.-K. & Hopfensitz, A. 2003. The co-evolution of individual behaviors and social institutions. *Journal of Theoretical Biology,* 223**,** 135-147.

Bowles, S. & Hammerstein, P. 2003 Does Market Theory Apply to Biology? 153-165. in *Genetic and cultural evolution of cooperation.* edited by P. Hammerstein. Cambridge, MA: MIT Press.

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.

Clarke, E. H. 1971. Multipart pricing of public goods. *Public Choice,* 11**,** 17-33.

Clutton-Brock, T. H. 1989. Review lecture: mammalian mating systems. *Proceedings of the Royal Society B,* 236**,** 339-372.

Clutton-Brock, T. H. 1991 *The Evolution of Parental Care,* Princeton, NJ: Princeton University Press.

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. *Philosophical Transactions of the Royal Society B,* 364**,** 807-819.

Dekel, E., Ely, J. C. & Yılankaya, O. 2007. Evolution of preferences. *Review of Economic Studies,* 74**,** 685-704.

Emlen, S. T. & Oring, L. W. 1977. Ecology, sexual selection and the evolution of mating systems. *Science,* 197**,** 215-223.

Enquist, M. 1985. Communication during aggresive 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.

Fudenberg, D., Kreps, D. M. & Maskin, E. S. 1990. Repeated games with long-run and short-run players. *Review of Economic Studies,* 57**,** 555-573.

Fudenberg, D. & Levine, D. K. 1992. Maintaining a reputation when strategies are imperfectly observed. *Review of Economic Studies,* 59**,** 561-579.

Gardner, A. & Grafen, A. 2009. Capturing the superorganism: a formal theory of group adaptation. *Journal of Evolutionary Biology,* 22**,** 659-71.

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. *Philosophical Transactions of the Royal Society B,* 355**,** 1581 -- 1591.

Grafen, A. 1990. Biological signals as handicaps. *Journal of Theoretical Biology,* 144**,** 517 -- 546.

Greif, A. 2006 *Institutions and the Path to the Modern Economy: Lessons from Medieval Trade,* Cambridge: Cambridge University Press.

Greif, A., Milgrom, P. & Weingast, B. R. 1994. Coordination, committment and enforcement: the case of the merchant guild. *The Journal of Political Economy,* 102**,** 745-776.

Grodzinski, U., Erev, I. & Lotem, A. 2007. Can hungry nestlings be trained to reduce their begging? *Behavioral Ecology,* 19**,** 116-125.

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.

Groves, T. & Ledyard, J. 1977. Optimal allocation of public goods: a solution to the "free-rider" problem. *Econometrica,* 45**,** 783-809.

Güth, W. 1995. An evolutionary approach to explaining cooperative behavior by reciprocal incentives. *International Journal of Game Theory,* 24**,** 323-344.

Güth, W. & Kliemt, H. 2000. Evolutionarily stable co-operative commitments. *Theory and Decision,* 49**,** 197-221.

Hauert, C. & Doebeli, M. 2004. Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature,* 428**,** 643-646.

Hoelzer, G. A. 1989. The good parent process of sexual selection. *Animal Behavior,* 38**,** 1067-1078.

Johnstone, R. A. 1999. Signaling of need, sibling competition, and the cost of honesty. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 12644-12649.

Johnstone R. A. 2000. Models of reproductive skew: A review and synthesis. Ethology 106:5–26.

Johnstone, R. A. & Grafen, A. 1992. The continuous Sir Phillips Sidney game: A simple model of biological signaling. *Journal of Theoretical Biology,* 156**,** 215-234.

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.

Klemperer, P. 1999. Auction theory: a guide to the literature. *Journal of Economic Surveys,* 13**,** 227-286.

Kreps, D. M. & Wilson, R. 1982. Reputation and imperfect information. *Journal of Economic Theory,* 27**,** 253-279.

Lehmann, L. & Keller, L. 2006. The evolution of altruism and cooperation -- a general framework and a classification of models. *Journal of Evolutionary Biology,* 19**,** 1365 -- 1376.

List, C. 2004. Democracy in animal groups: a political science perspective. *Trends in Ecology & Evolution,* 19**,** 168-169.

Maynard Smith, J. & Harper, D. G. C. 1988. The evolution of aggression: can selection generate variability? *Philosophical Transactions of the Royal Society B,* 319**,** 557-570.

Maynard Smith, J. & Parker, G. A. 1976. The logic of asymmetric contest. *Animal Behaviour,* 24**,** 159-175.

Maynard Smith, J. & Price, G. R. 1973. The logic of animal conflict. *Nature,* 246**,** 15-18.

Mcfarland, D. & Houston, A. 1981 *Quantitative ethology: the state space approach,* Boston, MA: Pitman Advanced Pub. Program.

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. & Sartori, A. 2008. Strategic uncertainty as a cause of war. *Quarterly Journal of Political Science,* 3**,** 327-352.

Milgrom, P. M. & Weber, R. J. 1982. A theory of auctions and competitive bidding. *Econometrica,* 50**,** 1089-1122.

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.

Myerson, R. B. 1981. Optimal Auction Design. *Mathematics of Operations Research,* 6**,** 58-73.

Myerson, R. B. & Satterthwaite, M. A. 1983. Efficient mechanisms for bilateral trading. *Journal of Economic Theory,* 29**,** 265-281.

Nash, J. 1950. The bargaining problem. *Econometrica,* 18**,** 155-162.

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.

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

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.

Orians, G. H. 1969. On the evolution of mating systems in birds and mammals. *The American Naturalist,* 103**,** 589-603.

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.

Riley, J. G. & Samuelson, W. F. 1981. Optimal Auctions. *The American Economic Review,* 71**,** 381-392.

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.

Rodríguez-Gironés, M. A., Zúñiga, J. M. & Redondo, T. 2002. Feeding experience and relative size modify the begging strategies of nestlings. *Behavioral Ecology,* 13**,** 782-785.

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. 2014. Incentives in the Family I: The Family Firm, An Evolutionary/Economic Theory for Parent-Offspring Relations. in *Human Nature, Early Experience and the Environment of Evolutionary Adaptedness.* edited by D. Narvaez, J. Panksepp, A. Schore & T. Gleason. New York: Oxford University Press.

Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. 2004. The evolution of cooperation. *Quarterly Review of Biology,* 79**,** 135-160.

Stamps, J. A. & Krishnan, V. V. 2001. How territorial animals compete for division of space: a learning-based model with unequal competitors. *The American Naturalist,* 157**,** 154-169.

Taylor, P. D. & Day, T. 2004. Stability in negotiation games and the emergence of cooperation. *Proc. R. Soc. B,* 271**,** 669-674.

Vehrencamp, S. 1983. A model for the evolution of despotic versus egalitarian societies. *Anim. Behav.*, 31, 667–682.

Vickrey, W. 1961. Counterspeculation, auctions and compatitive sealed tenders. *The Journal of Finance,* 16**,** 8-37.

Weibull, J. & Salomonsson, M. 2006. Natural selection and social preferences. *Journal of Theoretical Biology,* 239**,** 79-92.

West, S. A., Griffin, A. S. & Gardner, A. 2007. Evolutionary explanations for cooperation. *Current Biology,* 17**,** R661 -- R672.

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.

Young, H. P. 2008 Social norms. in *The New Palgrave Dictionary of Economics, Second edition.* edited by S. N. Durlauf & L. E. Blume. London: Macmillan.

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