Biological institutions: The political science of animal cooperation

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

Version May 29, 2010

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 socio-political 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, 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 developed in political science that focuses on social, political, and economic institutions presents new avenues for fruitful interdisciplinary synergy between political science and biology.

Our call for cross-disciplinary work between political science and biology has an impressive precedent in the seminal work of Axelrod and Hamilton on the evolution of cooperation (Axelrod and Hamilton, 1981). More recently, Conradt and List (2009) initiated an interdisciplinary project on collective behavior in animal groups that has already started bearing

[1] National Institute for Mathematical and Biological Synthesis (NIMBioS)
e-mail: erol@nimbios.org
[2] Department of Biology, Stanford University
[3] Department of Political Science, Stanford University
[4] New York University School of Law
[5] Hoover Institution

fruit (see the recent special issue of the Philosophical Transactions of the Royal Society). We endorse these recent efforts, and believe that the scope for interdisciplinary collaboration between political science and biology is even wider.

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 -- 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 written 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.

Animals do not have formal institutions such as legislatures, courts and committees and the absence of language constrains the complexity of social organization that is possible. Yet despite this, animal interactions do have structures that are determined by evolved norms and behavioral rules, in other words, the natural history of an interaction. These norms and behavioral rules determine where and when parties interact, how long and how close they stay together, what kind of dependency they have on each other, and so on. The main thesis of this

paper is that the natural histories of animal interactions can be fruitfully studied as informal institutions from the perspective of institutional theory.

This shift in perspective opens up new questions: behavioral ecology traditionally takes the context in which animals behave as given, and aims to explain why animals evolve to choose particular strategies in a given interaction based on the fitness consequences of different strategies. The institutional perspective adds to this goal the question of how the rules and norms that structure the interaction evolves,. One of the basic questions in behavioral ecology is what function a behavior serves. Similarly, the first question for animal institutions needs to be what the function of an institution is. Institutions in human societies frequently (though not always) are thought of as enabling interacting parties to arrive at mutually beneficial outcomes in the face of conflicting interests. In particular, institutions allow *efficient* outcomes, i.e. those fully capturing the gains from exchange or cooperation or avoid a case where people fail to achieve outcomes that would make everybody better off.

As a biological example, take 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, 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 and commitment problems can lead to inefficient outcomes.

Commitment, coordination, and private information

The question of efficiency is not entirely new to biology: the large and growing literature on the evolution of cooperation focuses on a special case of this question. This literature mainly focuses on how cooperation can evolve in the face of conflicts of interest using simple games such as the prisoners' dilemma (see reviews by Sachs et al., 2004, Lehmann and Keller, 2006, West et al., 2007a), where the choice is simply to cooperate or not. This literature has been very successful in elucidating the different ways a simple conflict of interest can be overcome by evolution. However, there is more than one way for conflict of interests to thwart cooperation and efficiency; specifically, conflicts of interests can cause problems of commitment, coordination and private information; all of which require different kinds of solutions to overcome. Institutional theory focuses on how to address each of these issues, focusing on the development of social norms, rules, and other constraints on individual behaviors. We illustrate how these problems arise in different settings and are dealt with by institutional theory, and how the institutional perspective can be applied to biology. We then discuss in some detail one of the major theoretical tools of institutional theory, mechanism design, and how it can be applied to biology. We follow this with a discussion of the implications of the institutional perspective for the levels-of-selection debate and conclude.

Commitment

Commitment: mediaeval merchants in a coral reef

Mediaeval 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 a merchant could trust the ruler of a distant city to honor his property rights and enforce the law, which would be a prerequisite for the emergence of long-distance trade. Even though the ruler 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 cannot commit to not robbing any merchant. Expecting this, merchants would be reluctant to visit this city, which would depress the volume of trade below the level that maximizes the total profits to the merchant and ruler combined. In other words, the inability of the ruler to commit to upholding merchants' rights and contracts 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. Since the potential loss of the ruler 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.

Two features of 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 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 all merchants to coordinate in the face of transgressions, thereby allowing the ruler to commit to honor whatever deal was struck. 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). Regardless of the interpretation, the guild institution increases 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 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 studies 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. There are efficient equilibria in which 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, the institution of the Law Merchant resolves the commitment problem faced by traders by allowing

successive trading partners to observe the reputation of each others. 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).

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 and remove ectoparasites. 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 feed on 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 of how clients keep cleaners from cheating and feeding on healthy tissue has been studied in great detail. Different client species seem to 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). 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 so that there is reputation effects involved (called image-scoring, Bshary, 2002). The combination of retaliation strategies and reputation effects makes it in the cleaner's best interest not to cheat clients, and thus provides a solution to a commitment problem similar to that faced by a mediaeval ruler. 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 retaliation rules represent a collection of simple informal institutions that together give rise to a biological market (Noë and Hammerstein, 1995, Noë et al., 2001). The aggregate effect of these rules is that over time more cooperative fish will be preferred by clients and thus get more "business", which exerts selection to maintain cooperation in this system. Hence, the market institution allows mutually beneficial exchanges in the absence of a system of laws and courts.

Recognizing the cleaner fish mutualism as collection of institutions is more than a matter of semantics; it illuminates some important issues. In economics, institutional theory has elucidated the implicit assumptions in neo-classical market theory, such as the costless and effective enforcement of contracts and the availability of information about prices (North, 1991), all of which have important repercussions on the way a market operates. For example, in order for the cleaner fish market to work effectively, a cleaner should be able to signal and commit to a "price" for its services (e.g. how much healthy tissue it will consume per minute of cleaning). This cannot be automatically assumed, as illustrated by the subset of cleaners who signal clients their cooperativeness, but then go on to cheat and feed on healthy tissue (Bshary, 2002). Market choices can prevent or limit such transgressions under some conditions. 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 with mechanism that keeps track of the long-

term performance of a cleaner. The territorial clients, because of their long-term association with a single cleaner, are good candidates for this role. 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. (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. 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. This model predicts that the interactions between territorial clients and choosy clients should play an important role in maintaining cooperation in the cleaner fish system.

The theory of the firm and breeding systems

Another important area where commitment problems play an important role is the theory of the firm. 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 as opposed to achieving production by on-the-spot transactions in the marketplace? 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 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, more subtle types of costs. For example, specificity of production assets to each other (e.g. a supplier building a plant next to a manufacturer's factory) 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 of a strong and weak predator cooperating to hunt prey. This inability to commit to not take advantage of the other party 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. (1978) argue that efficient transactions can be achieved in such cases when one party owns both assets, instead of relying on repeated market transactions.

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, 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. Hence, optimal investment decisions will be taken beforehand.

On 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, 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 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 polygamy (Clutton-Brock, 1991). The question is then when females can impose monogamy on males, either directly by choosiness, of indirectly through their distributions in space (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 of how the relationships between the mates should be organized. 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. However, 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 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 idiosyncratic in most bird species, whereas they are far less relationship-specific in mammals.

Under these crudely generalized conditions, 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. This argument suggests that transitions between monogamy and polygamy is possible even if the resource requirements and distributions of females stays the same (Clutton-Brock, 1989, Emlen and Oring, 1977). The theory-of-firm hypothesis predicts 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.

The theory-of-the-firm hypothesis for monogamy 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

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 (1988) investigate whether the coordination problem between two players is best resolved through a committee where players bargain, or through a market mechanism 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 creating the possibility 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 (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.

A biological counterpart to this 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's (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 randomize the decision. 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. 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

Private information in social interactions can hinder efficient outcomes because parties that have such information will often have incentives to not reveal it truthfully, even in cases of completely concordant interest (see "*Voting and information aggregation"* below). We will first illustrate private information problems in political science and biology using the issue of conflicts between states and animals.

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 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 strengths and value it places upon the object of contention, no feasible negotiated outcome may exist that looks preferable to both states before the war. 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 institutions that give states incentives to credibly reveal their private information without actually fighting need to be set up to prevent wars. 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 such institutions that enable credible revelation through "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. A similar result had been obtained earlier the related setting of bilateral

bargaining between buyers and sellers with private information (Myerson and Satterthwaite, 1983). 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).

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 imperfect information can lead to escalated conflicts. In the work that followed, Enquist (1985) 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. The honesty of these signals is maintained 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 there are some differences. First, biological models tend to be restricted to simple setups such as the Hawk-Dove game, and usually do not allow negotiated partitions of the resource that is being fought as a result of the 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 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 conflict, as in 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

Mechanism design

Most of the work in game theory specifies a game structure, 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 achieving overall efficiency or maximizing a certain player's expected payoff. A mechanism in game theory describes the rules of the game; in other words, mechanisms are mathematical representations of an institution.

Note that "mechanism" here means something different than the term's customary meaning in biology. In biology, mechanism refers to the processes that bring about biological phenomena. In the case of behavior, these biological mechanisms are widely termed "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 reminder 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 a parent company dealing with subsidiaries in a conglomerate, Groves, 1973), while others adopt a "social designer" perspective that is interested in improving social welfare (for example, authors of a constitution). A focal issue in much of mechanism design is to find "incentive compatible" mechanisms, where payoffs are structured as to make it optimal for players to reveal their information truthfully.

When used from a social planner perspective, mechanism design is a powerful 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, similar to how the concept of an evolutionarily stable strategy (ESS) is employed in behavioral ecology. The ESS approach is used to predict how animals should behave if they evolved to behave optimally in a strategic interaction, which provides adaptive hypotheses for behaviors observed in nature. The main difference is that instead of asking how an individual behaves, mechanism design would apply to the structure of the game and what effects it has on the overall outcome. On the other hand, similar to the ESS approach, mechanism design would not completely answer the question of how these outcomes are actually achieved (see "Evolutionary mechanism design" below).

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

(e.g., bargaining problems, or the design of a tax system) it is impossible to design a mechanism that will achieve first-best outcomes (e.g. ensure bilateral trade in every instance where it is efficient (Myerson and Satterthwaite, 1983) or balance the budget for the production of public goods (Groves and Ledyard, 1977)). Thus, 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 case of incentive compatible mechanisms can be found in a 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 fulfills the role of transfers in a mechanism design model. 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). A main difference between these two types of models is that in models of mate quality signaling, the signal costs are assumed to be determined by the choice of the signal (i.e. the length of a peacock's tail determines how costly it is), while in models of signaling need, the cost function is chosen by the receiver, i.e. the parent, to make truth-telling incentive compatible (Godfray, 1991, Nöldeke and Samuelson, 1999). In economic theory, the term

"signaling models" is traditionally used for the first type, whereas the second type of models is termed "screening models". The former assume that the informed player (e.g. the male) moves first by choosing a signal, while in the latter, the uninformed player (e.g. the parent) moves first by committing to a schedule of rewards as a function of the different signals the informed player can send.

In both types of models, the cost of signaling are wasted; they do not translate into fitness to either party (unless the "signal" is a cooperative act, such as providing a public good, see Gintis et al., 2001). 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. 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, Johnstone, 1999). This happens when variation in the quantity to be signaled is restricted, and the benefits from signaling (determined by average costs and benefits) is low relative to the costs that ensure incentive compatibility (determined by the marginal costs and benefits, Godfray and Johnstone, 2000).

What can a mechanism design perspective contribute to the biological theory of costly signaling? First, the mechanism design literature contains a variety of different approaches to solving incentive compatibility problems arising from private information with different background assumptions. For example, most costly signaling models in biology assume that individuals pay signal costs regardless of the eventual outcome of the game. This corresponds to an all-pay auction in mechanism design theory. An alternative assumption is that signaling takes place repeatedly, allowing individuals to learn the costs of receiving each item. At equilibrium then, individuals can choose signal signal only when it is worthwhile to do so; effectively giving

rise to a winner-pays mechanism (E. Akçay, unpublished results). Recently, Roughgarden and Song (in preparation) have modeled parental provisioning strategies using a model that incorporates a market-like process within the brood, similar to those that might operate within a conglomerate (Groves and Radner, 1972, Groves, 1973). These different assumptions give rise to qualitatively different set of predictions. For example, while the all-pay auction predicts signals not to be related to the particular realizations of the need of other individuals (because signals are chosen and costs are paid before other players reveal their information), the winner-pay and market-like mechanisms predict that they will be. In addition, the market-like mechanism of Roughgarden and Song predicts that the signal costs per food that are imposed by the parent will be a function of the realized need of the brood, whereas neither the all-pay and winner-pay auctions make this prediction. Thus, extending the existing signaling models with a mechanism design perspective helps generate alternative predictions for how animals should signal.

Second, most signaling models deal with restricted setups, usually with one-way signaling (one-to-one, or many-to-one). The complexity of the models increases rapidly when many individuals interact and signal to each other. In these cases, a foundational result of mechanism design theory, called the revelation principle (Myerson, 1979), can be extremely helpful. The revelation principle states that one can represent all (Bayesian) equilibria of a game with incomplete information with a special class of games, called direct mechanisms. In a direct mechanism, players simply announce their types (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). The revelation principle states that if an outcome can be implemented in some game, it can be also implemented by a direct mechanism. This allows one to ask whether an outcome is achievable, the magnitude of transfers and costs that are needed to

achieve an efficient outcome, or to characterize the best feasible outcomes while considering direct mechanisms only, greatly simplifying the problem. This principle has been applied to private information problems in the context of international conflict (e.g. Meirowitz and Ramsay, 2009, see above) and collective decision making (Meirowitz, 2006). In this way, the revelation principle can allow making comparative predictions without considering the full 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 resemblance to a recent model of collective decision making with communication and conflicting interests (Meirowitz, 2006).

Evolutionary mechanism design: proximate causes

Even though mechanism design theory can be a powerful tool to predict the outcomes of interactions trough strategic considerations, it does not completely answer how 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, or 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 work with very simple models of proximate

causation, where the behavior of an individual is determined by a single locus in a haploid genetic model. Increasingly however, both fields have been moving away from their traditional models. Models of bounded rationality and cultural influences on preferences (see Gintis, 2007 for an overview) are now being used widely in economics, 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 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 (representing the reward sensation of the agents) and act myopically to maximize these objective functions. These objective functions assume a role similar to a utility function in a boundedly-rational actor model. Accordingly, the behavioral dynamics of Akçay et al. (2009) result in a pure strategy Nash equilibrium of a game defined by the objective functions (which might be different than their material payoff 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 through the Merchant Guild (Greif et al., 1994) or the proposed institution in the cleaner fish system, both of which rely 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).

These models illustrate that the same outcome (cooperation) can be implemented through either through social institutions, or through internalized norms. Furthermore, these two methods can operate together in the same system. As an example, a strong argument has been made that human cooperation is maintained through punishment of cheaters, which makes cheating materially unprofitable. This corresponds to an institution. Yet punishing can be materially costly and individually suboptimal; 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 an institution providing material incentives or threats, and internal commitments. We believe that empirically motivated models of proximate causation need to be used to complement methods of mechanism design by asking how incentive compatible outcomes can be implemented through a combination of internal commitments 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. the dynamics of the rules of the game over time. Similarly, evolutionary game theory models in biology mostly assumes 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 revision). 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 pure conflict (as in the prisoners' dilemma) to mostly coordination (the hawk-dove game) can co-exist in a population.

Institutions and levels of selection

The main theme of this paper is that institutions can organize human interactions in ways that make self-interested parties achieve socially desirable outcomes, and that many biological interactions may be organized in similar ways to achieve cooperation, or efficiency. While the idea of an institution is a theoretical abstraction, it can be said that biological interactions are encoded in what is traditionally considered as the natural history of interactions, that is, the

description of who interacts with whom, when, how, and under which conditions. Hence, adopting the institutional perspective shifts attention from considering individual behaviors in the context of a given natural history to considering the evolution of natural history itself.

An important implication of this perspective is that the organization of biological interaction 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, but still ongoing, debate in evolutionary biology (Okasha, 2006) about whether selection at the level of groups is an important agent of evolution (Wilson and Wilson, 2007, West et al., 2007b), and whether social adaptations can sometimes be understood as group adaptations (Gardner and Grafen, 2009).

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 many 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 across species within the reef. Interdependencies between individuals' actions and benefits may create evolutionary pressures to resolve problems of commitment, coordination and information exchange both within and across species.

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. 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 very fruitful in leading to new questions and answers in our understanding of biological organization.

Acknowledgements

This paper was inspired by a workshop held in May 2009 at Stanford University, and funded by the Woods Institute for the Environment. We thank the participants of the workshop for discussion on the ideas that are advanced in this paper, and David Laitin and Ebru Erdem for comments on the manuscript. Any errors and omissions are ours. 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. (2007) Negotiation of Mutualism: Rhizobia and legumes. *Proc. R. Soc. B,* 274**,** 25 -- 32.

Akçay, E. & Roughgarden, J. (in revision) The evolution of payoff matrices: providing incentives to cooperate.

Akçay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. (2009) The evolution of other-regard integrating proximate and ultimate perspectives. *PNAS,* 106**,** 19061-19066.

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.

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.

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. & 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. *J Evol Biol,* 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.

Gintis, H., Smith, E. A. & Bowles, S. (2001) Costly signaling and cooperation. *Journal of Theoretical Biology,* 213**,** 103-119.

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, committment 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.

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

Groves, T. & Radner, R. (1972) Allocation of resources in a team. *Journal of Economic Theory,* 4**,** 415-441.

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.

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.

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.

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

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.

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. *TREE,* 19**,** 168-169.

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.

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 if 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.

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

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.

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.

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.

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

West, S. A., Griffin, A. S. & Gardner, A. (2007b) Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *J Evol Biol,* 20**,** 415-32.

Williamson, O. E. (1979) Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics,* 22**,** 233-261.

Wilson, D. S. & Wilson, E. O. (2007) Rethinking the theoretical foundation of sociobiology. *Q Rev Biol,* 82**,** 327-48.

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.