1

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]

Version August 25, 2011

[1] Corresponding author.

National Institute for Mathematical and Biological Synthesis (NIMBioS), University of Tennessee, Knoxville

e-mail: erol@nimbios.org

1534 White Avenue, Knoxville, TN, 37996

phone: (865) 974-9448

fax: (865) 974-9461

Current address: Department of Integrative Biology, UC Berkeley, Valley Life Sciences Building Berkeley, CA 94720

[2] Department of Biology, Stanford University

[3] Department of Political Science, Stanford University

[4] New York University School of Law

[5] Hoover Institution

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.

Keywords: commitment, coordination, private information, mechanism design, game theory

25

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

47   study questions in the other. There is a distinguished precedent for such collaboration, as

48   exemplified by the work of Axelrod and Hamilton (Axelrod and Hamilton, 1981) and the recent

49   focus on collective behavior in humans and animals (Conradt and List, 2009).

50   We call for a shift in perspective on animal behavior. Studies in behavioral ecology

51   traditionally take the context in which animals behave as given, and aim to explain why animals

52   evolve to behave in a particular way based on the fitness consequences of different strategies. In

53   contrast, the institutional perspective we propose focuses on how the rules and norms that

54   structure the interaction evolve.

55   In keeping with the behavioral ecology tradition that seeks to explain behaviors based on

56   their adaptive value, the first question in understanding the structure of an interaction is to ask

57   what function that structure serves. Political scientists have shown that formal and informal

58   institutions can have different functions: for example, they might allow certain parties (e.g. the

59   ruling elites of a country) to exert coercive control over other individuals. In many cases though,

60   institutions allow *efficiency*, i.e. fully capturing the gains from a social interaction to avoid

61   "leaving money on the table".

62   To illustrate the concept of efficiency with a biological example, consider two predators

63   sharing a prey killed. An outcome is efficient (from the predators' perspective) if the prey is

64   killed, and all the meat gets eaten. Killing the prey requires some measure of cooperation

65   between the two predators. Even though both predators would prefer the prey being killed and

66   eaten, but they still have conflicting interests between different efficient outcomes – who eats

67   how much – and their conflicting interests may cause them to fail to cooperate. For instance, if

68   the weaker of the two predators expects to be excluded from sharing after the prey is killed, it

69   might withhold its help during the hunt. Humans face such problems constantly, and an

70   incredibly diverse range of social institutions have arisen in myriad contexts to enable the human

71   equivalent of the stronger predator in this example to commit itself to share the kill.

72   At the heart of our proposal is the hypothesis that the structure of many biological

73   interactions may have evolved so as to sustain mutual benefit. This hypothesis is distinct from

74   group- and kin-selection approaches to the evolution of cooperation, since it does not depend on

75   variation among groups, or relatedness between individuals. Instead, it focuses on the

76   *organization* of the interaction, defined as who interacts with whom, when, and through which

77   proximate mechanisms, and how these features of an interaction may provide direct incentives

78   (i.e., selection through direct fitness benefits) that favor efficient outcomes in an interaction.[6]

79   This in turn leads to the question of how the organization of an interaction evolves as a result of

80   selection acting on the interacting individuals, and if and when such selection can lead to mutual

81   benefit. These questions remain almost completely unexplored in biology and are what we aim to

82   bring into focus with this review.

83   **Commitment, coordination, and private information**

84   The biological literature on the evolution of cooperation has largely focused on games

85   with simple conflicts of interest, such as the prisoners' dilemma or the snowdrift game (Axelrod

86   and Hamilton, 1981, Sachs et al., 2004, Hauert and Doebeli, 2004, Lehmann and Keller, 2006,

87   West et al., 2007). While these studies have elucidated the different ways a direct conflict of

88   interest can be overcome by evolution, other ways cooperation and efficiency can be thwarted

89   have received less attention. In this paper, we review three main strategic problems that can

90   preclude efficient cooperation, namely, the inability to commit, the need to coordinate and the

---

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

91   presence of private information. Each of these problems present distinct challenges, and require

92   distinct solutions.

93   Institutional theory from political science and economics provides a collection of

94   approaches to address these issues, focusing on the development of social norms and rules that

95   impose constraints on individual behaviors or shape the incentives individuals face. In this

96   review, we illustrate how institutional theory deals with commitment, coordination, and private

97   information problems, and how the institutional perspective can be applied to biology. We then

98   discuss in some detail one of the major tools of institutional theory, called mechanism design,

99   and discuss how an evolutionary mechanism design theory can be developed for biology.

100 Finally, we turn to the role of proximate mechanisms of behavior in structuring social

101 interactions, and the potential role of the institutional perspective in the resolution of the levels-

102 of-selection debate.

103   Before moving on, a note on terminology is in order. We use the terms "institution", or

104 "institutional perspective" in the biological context to refer to the evolved "rules of the game" of

105 an interaction. Just as the term "strategy" is used in biology refer to alternative behaviors with

106 different fitness consequences without implying conscious deliberation, our use of the term

107 "institution" is not meant to signify any formal organizational structure. Similarly, we use the

108 term "norm" not to refer to any normative prescription (e.g., one *ought* not to steal), but in the

109 accepted game-theoretic sense to refer to patterns of behavior that are self-enforcing (Young,

110 2008), either directly (e.g., cooperating due to others reciprocating) or through more complex

111 enforcement mechanisms (e.g., costly punishment of defectors).

112 **Commitment**

113 **Mediaeval merchants in a coral reef**

114   Mediaeval Europe was a tough place to do business, especially over long distances

115 (Greif, 2006). The rule of law was by no means assured, and property rights and the enforcement

116 of contracts could not be guaranteed by agreements between different polities as they are in

117 today's world. Despite these impediments, mediaeval trade somehow flourished. Greif et al.

118 (1994) tackle this puzzle with a stylized model that asks how a merchant could trust the ruler of a

119 distant city to honor the merchant's property rights and enforce the law. The basic dilemma for

120 the ruler is that even though he benefits from trade (through taxation), he can also rob any

121 merchant that comes to his city. Since any single merchant is unlikely to have great retaliation

122 power, the ruler would do best to rob any given merchant. Expecting this, merchants would be

123 reluctant to visit this city, which would depress the volume of trade, hurting both the merchant

124 and ruler. Thus, the inability of the ruler to commit to enforcing contracts and upholding

125 merchants' rights precludes efficiency.

126   Greif et al. argue that mediaeval merchants solved this problem by organizing themselves

127 into merchant guilds, which could declare bans on trade with a ruler that cheated one of its

128 members. Thus, if a ruler cheated a member of the guild, he would face retaliation from the

129 whole guild and not a single merchant, so that it becomes in the ruler's own interest not to cheat

130 *any* merchant belonging to the guild. Consequently, merchants belonging to the guild can trust

131 his promise not to do so; the commitment problem is solved.

132   Two points about this model are worth noting: first, the primary function of the merchant

133 guild is not to improve the bargaining position of the merchants to demand, say, a lower tax rate

134 from the ruler (although subsequently, the guild can do that as well). Rather, the guild enables

135 merchants to coordinate in the face of transgressions, which in turn makes it in the ruler's best

136 interest to honor property rights and agreed upon tax rates. Put another way, the merchant guild

137 replaces the collection of small, individual merchants that do not have retaliation power, and

138 hence act as short-run players, with a single, long-run player that does have retaliation power

139 (Fudenberg et al., 1990). In this way, the guild institution increases the payoff to both the

140 merchants and rulers. Second, while the merchant guild institution solves the ruler's commitment

141 problem, it creates another one, namely that of merchants committing to honor bans imposed by

142 the guild. Thus, the strategic problems faced by the merchants are fundamentally altered through

143 the formation of the guild institution. Greif et al. (1994) argue that complex rules and

144 interrelationships within guilds created incentives for merchants to honor the decisions made by

145 the guild.

146   A related model by Milgrom et al. (1990) studies the emergence of the Law Merchant to

147 resolve trade disputes in the Mediaeval age. In this model, a given pair of traders interact only

148 once with each other, but have the option of reporting transgressions to a private judge, who

149 keeps records of transgressions that are not remedied. There are efficient equilibria of this game

150 in which all traders consult (by paying a fee) the private judge before trade about whether their

151 prospective partners have outstanding judgments against them, withhold cooperation from those

152 who do, and report any transgression to the private judge after the trade. This equilibrium

153 sustains cooperation, because a trader who cheats a partner loses all future business, even though

154 he will never interact with that particular partner again. Note that the private judges do not have

155 any means of enforcement for their judgments. Nonetheless, the institution of the Law Merchant

156 ensures commitment to good behavior by allowing successive trading partners to observe each

157   other's past behaviors. In repeated games, reputation is one mechanism that allows long-run

158   players to commit to a given course of action (Kreps and Wilson, 1982, Fudenberg and Levine,

159   1992).

160   To see how similar ideas can be applied to a biological system, consider the cleaner

161   wrasses *Labroides dimidiatus* and *L. phthyrophagus* that live in coral reefs throughout the Indian

162   and Pacific Oceans. These fish occupy small territories, called cleaning stations, and are visited

163   by other fish (clients), which they inspect for ectoparasites and clean. In exchange for their

164   service, cleaners not only consume the ectoparasites, but they can also feed on the healthy mucus

165   and scales of their clients. Since consuming healthy tissue hurts the clients, but is preferred by

166   the cleaners, there is a conflict of interest between the partners in the mutualism (Bshary and

167   Grutter, 2002).

168   There is substantial research on how clients keep cleaners from cheating and feeding on

169   healthy tissue. Different client species seem to have different options available to them: some are

170   predatory and can in principle eat a cheater cleaner fish while others can simply choose to

171   associate with more cooperative cleaners (Bshary and Schäffer, 2002). Yet another class of

172   clients resorts to a punishment strategy, chasing the cleaner after being cheated (Bshary and

173   Grutter, 2002). Clients also observe other clients interacting with their prospective cleaner so that

174   there is reputation effects involved (Bshary, 2002). The combination of retaliation strategies and

175   reputation effects makes it in the cleaner's best interest not to cheat clients. Remarkably, cleaners

176   put in novel situations in laboratory experiments are able to learn new retaliation rules, and

177   adjust their behavior in order to optimize their gains (Bshary and Grutter, 2005, Bshary and

178   Grutter, 2006).

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

180 a biological market (Noë and Hammerstein, 1995, Noë et al., 2001, Bowles and Hammerstein,

181 2003). The aggregate effect of these rules is that over time the more cooperative cleaners will be

182 preferred by clients and thus receive more "business", which can maintain cooperation in this

183 system. Hence, the market institution allows mutually beneficial exchanges in the absence of a

184 system of laws and courts.

185   However, as economists have known for a while, markets can be inefficient in the

186 absence of institutions that provide for enforcement of contracts and the availability of reliable

187 information about the quality of service and prices (North, 1991). Similarly, for the cleaner fish

188 market to work efficiently, clients should be able to reliably assess the "price" (e.g. healthy tissue

189 eaten per parasite removed) each cleaner charges and cleaners need to be able to commit to their

190 "advertised" prices. These two features cannot be automatically assumed, as illustrated by the

191 presence of cleaners who signal clients their cooperativeness, but then go on to feed on healthy

192 tissue (Bshary, 2002). Market choices can prevent or limit such transgressions under some

193 conditions. However, when cleaners are saturated, the cost of losing any single client will be

194 small, and hence cleaners will not be able to commit not cheating any particular client, similar to

195 the rulers of mediaeval cities (Greif et al., 1994). This will depress visitation rates by clients; the

196 market will not function with full efficiently. Furthermore, a client fish will be limited to

197 information that it can directly gather by interacting with a cleaner or observing other clients'

198 interactions. Obtaining information in this way is likely to be costly and easily manipulated,

199 which also will limit the efficiency of a market.

200   An institution that keeps track of the long-term performance of a cleaner can solve these

201 problems. In particular, one might conjecture that some small territorial clients, who are long-

202 term associated with a single cleaning station, can serve as repositories of the cleaner's past

203   performance. If a cleaner cheats one of its prior clients, its territorial clients can make this known

204   to all future clients, which would then do best to avoid the cheating cleaner, similar to the Law

205   Merchant model (Milgrom et al., 1990). In this way, the cost of cheating for the cleaner is raised

206   from the marginal cost of losing one client to the cost of losing many or all clients, which makes

207   it in the cleaner's best interest to not cheat and solves the commitment problem. This equilibrium

208   also changes the incentives between the territorial clients and cleaners, and would prompt

209   cleaners to be cooperative towards their territorial clients. Empirically, such an arrangement

210   would predict that the interactions between territorial and roaming clients should play an

211   important role in maintaining cooperation in the cleaner fish system.

212   Note again that the argument above focuses on the structure of the game. It considers

213   how the incentives that the cleaner fish and its clients face change when a third party is

214   introduced, and whether this can lead to a higher level of efficiency at the level of the interaction.

215   In other words, it is about the evolution of games organisms play, rather than the strategies

216   within a given game. Such analyses are the hallmark of institutional theory in political science,

217   but are rarely done in biology (but see Worden and Levin, 2007, Akçay and Roughgarden, 2011)

218   **The theory of the firm and breeding systems**

219   An important area of economics where commitment problems play a central role is the

220   theory of the firm. A firm in economics is defined as an organization that produces goods and

221   services outside the marketplace, by means of contracts that last much longer than each action

222   the agents take (e.g. producing a single unit of goods). These contracts frequently concentrate the

223   ownership of production assets and decision-making in some agents (the employer) and

224   remunerate others (the employees) in return. One of the main questions in the theory of the firm

225   is why firms exist at all, i.e. why do agents organize themselves into long-term relationships as

226 opposed to achieving production by on-the-spot transactions in the marketplace? This question,

227 first posed by Ronald Coase in his influential essay (Coase, 1937), has stimulated a large body of

228 research in law, economics, and political science.

229   Coase proposed that firms form to reduce the costs that arise from making repeated

230 transactions in the marketplace, such as the cost of finding out the market price of goods and

231 services and negotiating over the terms of agreement. A more subtle type of cost occurs when

232 production assets are specific to each other (e.g. a supplier needs to build a plant next to a

233 manufacturer's factory). In such cases, recourse through partner choice in the open market may

234 become infeasible. In particular, situations that are not covered in the original agreement

235 between parties (e.g. unexpected movements in prices) would create incentives for the

236 advantaged party to try to appropriate gains (or avoid losses) from those situations, to the

237 detriment of the other party – much as in the example of a strong and weak predator cooperating

238 to hunt prey. The inability of parties to commit to not taking advantage of each other can then

239 lead to underinvestment into the production assets relative to what is efficient, because both

240 parties expect that their sunk costs from the specific investments will be appropriated by the

241 other. Klein et al. (1978) argue that efficient transactions can be achieved in such cases when

242 one party owns both assets, instead of relying on repeated market transactions. Likewise,

243 Grossman and Hart (1986) show that if investments are relationship-specific and contracts

244 cannot account for all possible contingencies, the problem of underinvestment can be solved by

245 assigning control of both assets to one party (the firm), which removes the incentive to

246 appropriate sunk costs after the investments are made. In a related vein, Williamson (1979)

247 argues that long-term repeated transactions with highly specific or idiosyncratic assets should be

248 governed by what he calls relational contracts, which specifies the roles of two parties in an

249 ongoing relationship (e.g. partners, or employees and employers), rather than particular actions

250 they need to take in each possible state of the world.

251   How can these ideas be applied in biology? Consider the striking contrast between the

252 breeding systems of birds, where the overwhelming majority of species exhibit social monogamy

253 (i.e. a male and a female raising offspring together), versus those of mammals, where most

254 species are polygamous. Previous thinking on this pattern mostly departs from the assumption

255 that females prefer monogamy, and males polygyny (Clutton-Brock, 1991), and asks when males

256 are constrained to be monogamous, either through female choice, or due to females' distribution

257 in space and time (Orians, 1969, Emlen and Oring, 1977, Clutton-Brock, 1989).

258   Viewing offspring rearing as analogous to the production of goods by a firm leads to a

259 new perspective that predicts not only the occurrence of monogamy versus polygamy, but also

260 the nature of the relationships between the mates. In mammals, lactation implies that females

261 control offspring provisioning, so that males cannot directly invest into that component of care.

262 However, males can invest indirectly into provisioning through feeding the female (or allowing

263 her to forage undisturbed), and also into other components of care, such as predator protection.

264 Roughgarden (2009) proposed that polygyny may evolve because males compensate for being

265 excluded from making investments directly into the offspring.

266   An alternative hypothesis can be constructed by observing that most mammals, females

267 are mobile during gestation and offspring are either mobile shortly after birth, or can be carried

268 around. Hence, mammal females can receive help from different males, for example, by moving

269 between territories or associating with different males, without necessarily losing their offspring

270 (especially if they have copulated with multiple males, Wolff and Macdonald, 2004). A female

271 bird with altricial young, however, cannot move her eggs or nestlings to another nest; likewise,

272   the male cannot share the nest and the territory he has invested in with another female while

273   another brood is in it. Hence, investments by the male and female into breeding are relationship-

274   specific in most bird species, whereas they are far less so in mammals.

275   Under these crudely generalized conditions, the theory of the firm predicts that

276   interactions between mates in birds should be governed by "relational contracts", i.e. longer-term

277   commitments. The pair-bond of birds would be analogous to a relational contract, if it

278   proximately alters individuals' behavioral tendencies to act outside their short-term selfish

279   interest (Akçay et al., 2009) and according to the needs of the brood (Akçay and Roughgarden,

280   2009). Of course, for such proximate commitment mechanisms to evolve, parties need to be able

281   to reliably ascertain each other's commitment before the breeding attempt. Extensive courtship

282   behaviors in birds can be hypothesized to serve this function (in addition to determining the

283   likely quality of parental care supplied by the male; Hoelzer, 1989). In contrast, interactions

284   between mates in mammals can be maintained by repeated shorter-term commitments (or on-the-

285   spot transactions) since each party maintains an outside option due to their less partner-specific

286   investment. This argument suggests that transitions between monogamy and polygamy is

287   possible even if the resource requirements and distributions of females stays the same (Emlen

288   and Oring, 1977, Clutton-Brock, 1989). The theory-of-firm hypothesis thus holds that

289   monogamy and social pair bond will evolve not because of the inability of males to monopolize

290   females, but because of the inability of the males and females to switch partners without losing

291   their past investments. This prediction can be tested by taking advantage of the variation in

292   offspring mobility in different mammal and bird species: it predicts that, regardless of home-

293   range sizes, species with more mobile offspring will be more likely to have social polygamy.

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

295 monogamy.

296 # Coordination

297 **How to agree on industrial standards**

298   Many social and economic situations are characterized by the need to coordinate to

299 achieve mutual benefit. Coordination problems arise in settings that range from competition

300 between different industrial standards to the movement decisions of an elephant herd. Even in

301 cases where the interests of the parties are largely concordant, coordination failure can preclude

302 efficiency.

303   In an influential paper, Farrell and Saloner (1988) compare two different institutions

304 through which two players can coordinate their actions. The first is a committee where players

305 negotiate with each other, while the second one is a market mechanism where both players come

306 forward with their own actions and hope that the other follows suit. The committee institution

307 ensures coordination, but imposes negotiation costs (in particular, delays in agreement), while

308 the market minimizes delay costs but creates the possibility of mis-coordination if players

309 commit to different actions at the same time. Farrell and Saloner show that a hybrid institution

310 that prescribes negotiations while also allowing players to individually commit to a course of

311 action at every possible stage does best compared to both of the pure institutions. More recently,

312 Farrell and Simcoe (2009) study when, in the context of deciding industrial standard, the use of

313 costly tactics to persuade the other party to give up (i.e., playing a war-of-attrition game) will be

314 optimal when two proponents have private information about the quality of their proposals. They

315 find that when there is no vested interest (i.e. no conflict over the eventual standard), the war-of-

316 attrition game chooses the best standard without delay, and thus achieves the most efficient

317 outcome. However, when the vested interest is high enough, it becomes optimal to employ an

318 institution that allows the war-of-attrition to proceed until a specified time and if the game is

319 unresolved at that time, chooses an outcome randomly. Interestingly, both players prefer such an

320 institution to the unchecked war-of-attrition before they learn their own proposals' quality, but

321 they might prefer continuing with the unchecked war-of-attrition game after they have learned

322 this information. This means that a mechanism needs to be in place to commit both sides to

323 accept the randomly chosen outcome in case of persistent disagreement.

324   A biological coordination problem can be found in collective decision-making with

325 multiple alternatives, an area that has seen a rapid growth in recent years (Conradt and Roper,

326 2005, Conradt and List, 2009). The results of Farrell and Saloner (1988) suggest that a

327 combination of consensus building through communication and individual initiative will be

328 optimal for cases with perfect information about the alternatives. This would predict that we

329 should observe a mix of consensus decisions and individual decisions even in cases where there

330 is no reason to expect one particular individual to be the decision maker. On the other hand,

331 Farrell and Simcoe (2009) suggest that when the decision is held up between two parties with

332 conflicting interests (e.g. seeking water vs. seeking food), it might be optimal for a third,

333 uninterested party, to randomize the decision. For example, the individual in best condition in the

334 group (e.g. the most satiated and hydrated) can arbitrarily make the decision, since such an

335 individual would represent the closest approximation to a neutral party in the group. In general,

336 collective decision in humans and animals making also presents problems of private information,

337 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.

360 War may then follow as a means of credibly revealing (or bluffing about) one's private

361 information. However, under some circumstances, practices in international "militarized

362 disputes" that are short of a full-blown war – such as mobilizing troops and issuing public

363 statements that have domestic consequences – may also enable credible revelation through

364 "costly signaling."

365   Another way that a full-blown war might be prevented is through international

366 institutions that give states incentives to truthfully reveal their private information. Recent

367 literature on the political science of war asks what is the best international institutions can do in

368 reducing the probability of war (e.g. Fey and Ramsay, 2009, Meirowitz and Sartori, 2008). One

369 important result from this literature is that when states' private information is correlated with

370 each other (e.g., when the information is about military capabilities) then regardless of the details

371 of the negotiation process between the states there is a positive risk of fighting, provided that the

372 total costs war are not too high. This is similar to the foundational result of Myerson and

373 Satterthwaite (Myerson and Satterthwaite, 1983) in the setting of bilateral trade between buyers

374 and sellers with private information, which states that there exist no trading scheme that

375 guarantees efficient trades to always happen as long as there is some probability of inefficient

376 trades. Both of these results are obtained using a powerful theorem from mechanism design

377 theory, called the revelation principle (Myerson, 1979, see also the section on mechanism design

378 below), and put limits on how efficient the outcome of an interaction can be made given that

379 there is private information.

380   Biologists have long used the metaphor of war and peace for agonistic behavior between

381 animals (Maynard Smith and Price, 1973), and there are some close parallels between previous

382 theory in animal behavior and political science. Maynard Smith and Parker (1976) were among

383 the first to recognize that private information can lead to escalated conflicts. Subsequent studies

384 (Enquist, 1985) (Maynard Smith and Harper, 1988) showed that when two animals have private

385 information about their valuation of the resource or their fighting ability, evolutionarily stable,

386 costless signals can exist and allow animals to avoid fighting some of the time. In this case, the

387 honesty of these signals is maintained not through intrinsic costs, but because individuals who

388 pretend to be stronger than they are face fights with stronger individuals. Individuals who

389 pretend to be weaker, on the other hand, experience opportunity costs due to the contests they

390 could win without fighting, but have to fight out because of their signals.

391   The focus in biology on how animals avoid fighting costs parallels the political science

392 literature on crisis bargaining and international institutions to prevent war, but with some

393 important differences. First, biological models tend to focus on simple games such as the Hawk-

394 Dove game, and usually do not allow negotiated partitions of the contested resource through pre-

395 fight interactions. In many interactions, however, the contested resource can be divided, such as

396 territorial interactions (Stamps and Krishnan, 2001, Pereira et al., 2003) or bargaining over

397 resource exchange (Akçay and Roughgarden, 2007). Bargaining models commonly used in the

398 theory of social and political institutions can therefore be used to extend existing biological

399 theory to cases where a near-continuous division of the resource is plausible. Furthermore, a

400 mechanism design approach can help to extend these results to generate "game-free" results

401 about the expected frequency and level of conflict (Fey and Ramsay, 2009).

402 **Voting and information aggregation**

403   How groups of individuals make decisions is one of the central questions in both political

404 science and behavioral ecology and has recently seen fruitful collaborations between the two

405 disciplines (List, 2004, Conradt and List, 2009). Voting theory is concerned with how and

406   whether efficient outcomes can be achieved when multiple individuals have to make a collective

407   decision. The social science literature usually takes the point of view of a "social designer",

408   whose goal is to satisfy group-level criteria, such as efficiency, for the aggregate decisions. In

409   contrast, biologists usually focus on individually optimal strategies in a given game (Conradt and

410   Roper, 2009).

411   An interesting problem in voting theory arises when individuals have private information

412   about their own preferences or the environment. Suppose for the moment that individuals share a

413   common interest but have independent estimates of what action best furthers the common

414   interest. If all individuals have the same and better-than-even-chance of being accurate, then

415   consulting more individuals and taking the average of all estimates, on average, improves the

416   accuracy of the decision. This result is known as the Condorcet Jury Theorem. Problems arise,

417   however, when group members decide strategically about whether or not they will reveal their

418   information truthfully, a behavior called strategic voting. Surprisingly, in such cases it may pay

419   for individuals to withhold or misrepresent their information, even with complete concordance of

420   interest (Austen-Smith and Banks, 1996, Austen-Smith and Feddersen, 2009). The reason is that

421   when deciding how to vote, a rational individual should consider only the case where he or she

422   would be casting the deciding vote; in all other cases it doesn't matter what he or she does. But

423   the event that the focal individual's vote is pivotal implies that the other individuals are voting in

424   a particular way (e.g. under unanimity rule, all others must be voting "Yea" for the focal

425   individual to be pivotal). Hence, the focal individual should condition her vote upon that state of

426   the world, which might make voting against one's private signal optimal. Therefore, strategic

427   voting is frequently not informative about the private information of agents. Generally speaking,

428   the incentives to misrepresent one's information are determined by how likely it is that a player

429   will cast the decisive vote in determining the outcome. Thus, these incentives are a more

430   significant problem in smaller groups, since each individual has a higher probability of being

431   pivotal. Conversely, in large groups the Condorcet Jury Theorem approximately survives

432   strategic voting (Feddersen and Pesendorfer, 1997).

433   Not surprisingly, the problem of strategic voting and misrepresentation is aggravated

434   when there are real conflicts of interests within the group. However, Meirowitz (2006) shows

435   that outside transfers to individuals as a function of their revealed information can create such

436   incentives, and that the magnitude of the required transfers becomes smaller as group size gets

437   larger (due to, again, each individual having smaller chance of being pivotal). These results have

438   not yet seen use in the biology of group behavior, but have connections to the costly signaling

439   theory in biology, as we discuss below.

440   **Mechanism design, auctions and costly signaling**

441   We now turn to mechanism design theory and the related theory of auctions.

442   Conventional game theory starts by specifying a game structure, and predicts outcomes

443   supposing self-interested agents with some level of computational capacity and access to

444   specified public and private information. Mechanism design inverts this approach: it specifies the

445   information structure and a set of possible games, and finds the game that produces the outcome

446   with a desired property, such as achieving efficiency. Auctions are particular instances of this

447   problem where the question is how to sell an object to a buyer in a way that maximizes the profit

448   to the seller.

449   The term "mechanism" in game theory is used to describe the "rules of the game"; in

450   other words, mechanisms are mathematical representations of an institution. In contrast,

451   biologists use the term "mechanism" to refer to the processes that bring about a behavior, i.e. the

452   "proximate cause" of behavior, such as the firing of specific neural circuits in the brain or

453   changes in hormone levels. To avoid confusion, we use the term "mechanism" by itself in the

454   game theoretic sense, and refer to biological mechanisms with the phrase "proximate

455   mechanism". But the two senses of the term mechanism are intimately connected: game-

456   theoretic mechanisms represent the structure of the interaction generated by proximate

457   mechanisms of behavior.

458   Mechanism design theory primarily deals with situations where agents have private

459   information and face incentives to not reveal this information accurately, as in the cases of

460   international conflict and voting discussed above. Some mechanism design models take up the

461   perspective of a player who has the power to alter the game structure to better serve its own

462   interest, for example a parent company dealing with subsidiaries in a conglomerate (Groves,

463   1973) or an auctioneer who tries to set up an auction that provides the most revenue from the sale

464   of an item (Vickrey, 1961, Klemperer, 1999). Other mechanism design problems presume a

465   "social designer" that is interested in improving aggregate welfare, for example, an international

466   organization trying to prevent wars (Fey and Ramsay, 2009) or a government setting up a tax

467   schedule to provide public goods (Clarke, 1971). Regardless of the objective of the designer, the

468   focal issue in mechanism design is to set up a game that makes truthful revelation of information

469   individually optimal, i.e. incentive compatible. Fundamentally, incentive compatibility is means

470   aligning the interests of a self-interested agent with whatever the objective of the mechanism is.

471   This frequently requires payments to be made to, or deducted from, the individuals as a function

472   of the messages they reveal.

473   Incentive compatible mechanisms are crucial to achieve cooperation in diverse situations

474   such as bargaining problems (Myerson and Satterthwaite, 1983) or producing public goods

475   (Groves and Ledyard, 1977), where the mutually beneficial policy will depend on the private

476   information of the individual agents. One general result from mechanism design is that ensuring

477   incentive compatibility frequently leads to ineluctable costs when there are conflicting interests

478   between agents (Myerson and Satterthwaite, 1983). The difference between the overall payoff

479   that could have been achieved if there was no private information and that achieved under an

480   incentive compatible mechanism is sometimes termed "agency loss".

481   In biology, private information problems have been investigated in three main contexts:

482   agonistic behavior (Maynard Smith and Parker, 1976, Enquist, 1985; see also "International

483   conflict and how to prevent it" above), signaling of mate quality (Zahavi, 1975, Grafen, 1990),

484   and signaling of need between parents and offspring or other relatives (Godfray, 1991, Johnstone

485   and Grafen, 1992, Nöldeke and Samuelson, 1999, Godfray and Johnstone, 2000). The method of

486   analysis in most of these models is to assume that there is a costly behavior or trait -- the signal --

487   that the informed party (e.g., males who know their quality, or offspring who know their need)

488   can invest in. The honest (or separating) signaling equilibrium has to satisfy two conditions: (1)

489   at the optimal signal level of the signaler given the cost of signaling and the benefit resulting

490   from the uninformed party's reaction to the signal, is different for different values of the private

491   information, (2) assuming that the signaler has signaled truthfully, the receiver reacts to the

492   signal to maximize its own payoff. Conditions (1) and (2) imply that incentive compatibility is

493   satisfied for the signalers and the receiver has gained full information. In all these models, costs

494   of signaling are wasted, reflecting agency loss. In fact, the agency loss in parent-offspring

495   signaling can be so high as to render one or both parties worse off relative to no signaling

496   (Rodríguez-Gironés et al., 1996, Johnstone, 1999, Godfray and Johnstone, 2000).

497   What can mechanism design theory do for biology? We see potential contributions of the

498   mechanism design approach in terms of methodology and increasing the scope of strategic issues

499   considered in models.

500   In terms of methodology, mechanism design provides powerful and well-developed

501   techniques to characterize and solve incentive compatibility problems. Chief among these

502   techniques is the revelation principle (Myerson, 1979); which states that any (Bayesian Nash)

503   equilibrium of any game with incomplete information can be represented as an equilibrium in a

504   special class of game called direct mechanisms. A direct mechanism is a game where all

505   individuals reveal their information truthfully to a (imagined) central arbiter, who then decides

506   what outcome of the game is implemented and which payoff each player will receive based on

507   the revealed information. To implement the payoffs, the direct mechanism may make payments

508   to the players or impose costs. The revelation principle allows one to answer many questions in

509   mechanism design, such as determining the best outcome that can be achieved under incentive

510   compatibility constraints, by focusing on direct mechanisms only which allows generating very

511   general results for a large class of games with imperfect information.

512   One such result obtained is the revenue equivalence theorem, which states that when

513   buyers in an auction have independent valuations about an object, many commonly used auction

514   mechanisms (such as the English and Dutch auctions) yield the same expected revenue to the

515   auctioneer (Vickrey, 1961, Myerson, 1981, Riley and Samuelson, 1981). To see why this is true,

516   we first note that all auctions can be represented by a direct mechanism, and in a direct

517   mechanism, incentive compatibility requires that the payments charged to each bidder have to

518   balance his incentives to misrepresent. These incentives depend on the bidder's valuation for the

519   object and the probability that he wins the object. Since the valuation is assumed to be

520   exogenous, the incentive compatible payments are only a function of the probabilities each

521   bidder wins the object (up to a boundary condition, Myerson, 1981, Milgrom and Weber, 1982).

522   Since many common auction mechanisms result in the same probabilities of winning for a bidder

523   (they all give the object to the bidder with the highest valuation) and have payment rules such

524   that the buyer with the lowest valuation has zero expected payoff, they all result in the same

525   expected payments. Armed with the revenue equivalence theorem, the optimal auction

526   mechanism that maximizes the proceeds for the auctioneer can be found among a very large

527   class of possible mechanisms (Myerson, 1981, Riley and Samuelson, 1981).

528   Results such as revenue equivalence can be applied to biology, as exemplified by the case

529   of multiple offspring signaling their need to the parent. Johnstone (1999) analyzed what amounts

530   to an all-pay auction where multiple offspring "bid" for a food item by engaging in costly

531   begging, and finds the signal costs that make truthful begging optimal. The revenue equivalence

532   result implies that the expected costs will be independent from the details of the signaling

533   mechanism as long as the item, if awarded, is given to the highest bidder. Hence, one can

534   analyze computationally simpler mechanisms to determine the expected costs. More importantly,

535   one can use an analysis similar to Myerson's (Myerson, 1981) to determine the optimal auction

536   rule that maximizes the profit to the parent (i.e., the auctioneer). The difference with the auction

537   models in economics is that an auctioneer wants the bidders to pay a higher cost (as this is his

538   income), while the costs in parent-offspring signaling are deducted from the parent's fitness.

539   Apart from the methodological contributions, mechanism design and auction theory also

540   suggest a number of new strategic problems that have biological significance. One such problem

541   that arises in auctions is when the bidders do not have entirely independent valuations of the

542   items offered. This so-called "common values problem" arises, for example, in auctions for oil

543 drilling leases, where the value of the lease to all players depends on how much oil there really is

544 beneath the soil, but different players might have different estimates for how much oil there is,

545 and may have different costs (due to technological or logistic reasons) for extracting and

546 processing it. A similar problem arises in biological signaling where the signalers do not know

547 the objective quality of the item offered; for example, offspring may not know how nutritious a

548 given food item is, or different males courting a female might have different estimates of the

549 breeding ability of a female. In these cases, the value of "winning" the auction to the signalers

550 will not be independent. In general, revenue equivalence breaks down for common value

551 auctions (Milgrom and Weber, 1982); that is, different auction mechanisms produce different

552 expected proceeds to the auctioneers. Furthermore, a common phenomenon associated with

553 common values auctions has been termed the "winner's curse", where the fact that a bidder has

554 won an auction on average means that he has overestimated the real value of the item. Hence,

555 optimal bidding in common values auctions usually involves individuals "shading" their bids, i.e.

556 bid lower than their actual valuations, which reduces the revenue to the auctioneer. Therefore,

557 the auctioneer can sometimes improve its expected payoff to reveal whatever information it has

558 truthfully (Milgrom and Weber, 1982). However, revenue in economic models corresponds to

559 wasted signal costs in biological signaling; which suggests, for example, that parents in a costly

560 begging setup might in fact do better to hide the real value of the food item from the offspring, so

561 as to decrease the incentives to engaging in excessively costly begging. This, and other strategic

562 complexities arising from common values remain to be explored.

563   Finally, an understudied issue in biological signaling is the role of behavioral dynamics in

564 determining the outcome of signaling interactions. Existing biological models of signaling for

565 the most part consider both the signaler and receiver behavior as evolving at the evolutionary

566   time-scale. In reality many of these interactions involve the parties reacting to each other in the

567   behavioral time-scale and are able to "learn" what different signals mean or what reactions they

568   elicit (Rodríguez-Gironés et al., 2002, Grodzinski et al., 2007). Two recent studies, by

569   Roughgarden and Song (in press), and Akçay (in review) model parent-offspring communication

570   in the behavioral time-scale. Roughgarden and Song consider a supply-demand dynamics

571   between the parent and the offspring, with the parent adjusting the "price" of food (i.e., the

572   amount of begging the parent requires for unit food delivery), and the offspring adjusting its

573   demand in response to the price it sees. Roughgarden and Song consider the "market-clearing"

574   equilibrium of this dynamic where the offspring demand exactly the amount of food the parents

575   are willing to provide, and show that this equilibrium is separating. Akçay, on the other hand,

576   models the communication process as a dynamic auction process, similar to the canonical model

577   of mechanism design, the Vickrey-Clarke-Groves auction (Vickrey, 1961, Clarke, 1971, Groves,

578   1973). Akçay assumes that the parent has a genetically determined decision rule specifying at

579   which rate food is yielded in response to offspring begging. The offspring "learn" to beg

580   optimally given the parent's decision rule, leading to a separating equilibrium. Akçay then

581   considers the evolution of the parent's decision rule, and shows that under some parameter

582   values, the parent can evolve to let the offspring demand food without any cost. The behavioral

583   equilibrium induced by this decision rule is still separating, since offspring have a partial interest

584   in the parent's survival, and therefore lower need offspring demand food at lower levels. In the

585   language of mechanism design theory, the parent's decision rule corresponds to a mechanism,

586   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

610 constraints on computational capabilities and cultural influences on preferences (Gintis, 2007),

611 while biologists have been developing game theoretic models with explicit mechanisms of

612 proximate causation (McNamara et al., 1999, Taylor and Day, 2004, Roughgarden et al., 2006,

613 Akçay et al., 2009, Roughgarden, 2009).

614   Of these recent models in biology, the one by Akçay et al (Akçay et al., 2009) provides

615 many immediate linkages to economic theory and a long line of thinking in behavioral ecology

616 that views animals as goal-oriented agents (McFarland and Houston, 1981). In this model,

617 individuals have a genetically encoded objective function (representing the reward sensation of

618 the agents) and act myopically to maximize these objective functions. These objective functions

619 assume a role akin to the utility function in a boundedly rational actor model. The behavioral

620 dynamics of Akçay et al. (2009) result in a pure strategy Nash equilibrium of a game defined by

621 the objective functions, which can take into account things other than the material payoffs of

622 individuals. The equilibrium actions of the individuals determine individuals' fitness, and result

623 in natural selection acting on the objective functions. Using this framework, Akçay et al. show

624 that other-regarding objective functions can evolve even in the absence of kin- or group-

625 selection, and under complete conflict of interest between the players. This result provides partial

626 support for the conjecture by Roughgarden et al. (Roughgarden et al., 2006, Roughgarden, 2009)

627 that social behaviors might be driven by non-competitive dynamics such as the "team-play

628 dynamics", which leads to the Nash Bargaining Solution (Nash, 1950) as a behavioral

629 equilibrium. The model of Akçay et al. is also related to "indirect evolution" models in

630 economics (e.g., Güth, 1995, Weibull and Salomonsson, 2006, Dekel et al., 2007), which

631 consider the evolution of individuals' preferences in social interactions.

632   Non-selfish objectives, or preferences, can be interpreted as internal commitments to

633   cooperate against the actor's short-term material interest (Güth and Kliemt, 2000). This

634   mechanism of commitment contrasts with the model of the Merchant Guild (Greif et al., 1994),

635   which relies on pure self-interest of the actors. These two different modes of commitment

636   require overcoming different challenges: the Merchant Guild has to be able to make the threat of

637   collective retaliation by the Guild credible, while the other-regarding individual has to be

638   sufficiently certain of the preferences of its partner in order to be not taken advantage of (Ok and

639   Vega-Redondo, 2001).

640   The significance of such internal commitments is that they change the incentives that the

641   social partners of the focal individual face. In Akçay et al's model (2009), other-regard evolves

642   because individuals that face an other-regarding individual become better off helping the focal

643   individual, which in turn increases the fitness of the focal individual. Similarly, in the costly

644   signaling models of parent-offspring communication (Godfray, 1991, Johnstone, 1999, Akçay, in

645   review), the proximate mechanisms that underlies the parent's allocation rule and the offspring's

646   decisions to signal determine how much the offspring beg for food and what costs they incur. As

647   such, we expect there to be direct selection on these proximate mechanism of decision-making,

648   through which the game theoretic mechanism ensuring truthful communication can evolve. More

649   recently, Akçay and Roughgarden (2011) consider how traits that provide explicit incentives to

650   others to cooperate can evolve, and find that evolution can lead to polymorphisms in the

651   incentives provided, which sustains much cooperation in the population, but also generates a

652   complex pattern of behavioral diversity.

653   Another example of proximate mechanisms changing incentives is the idea that

654   punishment of cheaters that can maintain cooperation by making cheating materially

655 unprofitable. Yet punishing can be materially costly and individually suboptimal; hence Gintis

656 and colleagues (Gintis et al., 2003, Gintis, 2003) argue that a commitment to punishing is

657 achieved through the evolution of other-regarding preferences. Thus, simple informal institutions

658 that provide material incentives and disincentives can be implemented through proximate

659 mechanisms of behavior determining individual decision-making.

660   Thus, an increasing number of studies investigate the idea that social interactions are

661 structured by evolution changing incentives that parties face in behavioral time-scales, but much

662 remains unexplored. We believe that this issue represents one of the major areas of development

663 for evolutionary biology. The institutional perspective, coupled with models that integrate

664 proximate mechanisms in evolutionary models hold great promise for tackling these questions.

665 **Institutions and levels of selection**

666   Finally, the institutional perspective we advocate in this paper puts a new twist to an

667 ancient, but still ongoing, debate in evolutionary biology about whether some social adaptations

668 can sometimes be understood as serving the "good of the group" (Okasha, 2006, Gardner and

669 Grafen, 2009). For example, consider a coral reef with many interacting species, including the

670 cleaner wrasse and its clients. Neither kin selection nor group selection can act on a coral reef

671 ecosystem as a whole, since species are unrelated to each other and there is no population of

672 coral reefs that are subject to differential mortality and reproduction as a whole. However, it is

673 possible for natural selection to shape the organization of the interactions in a coral reef though

674 acting on individual species' traits, and align selection at the individual level with the objective

675 of maximizing the mutual benefit at an aggregate level. To be sure, this is by no means an

676 inevitable outcome of natural selection, yet at the same time, it represents an understudied

677 possibility in evolutionary theory. The institutional perspective provides the theoretical

678   framework to investigate when and how efficiency at the aggregate level can be attained in the

679   face of conflicting interest, and problems of commitment, coordination and private information.

680   Recently, Bowles et al. (2003) modeled how group-level institutions in humans (such as

681   taxes to share benefits and intra-group assortment mechanisms) can evolve through group

682   selection (either cultural or biological). In their model, institutions evolve because they facilitate

683   the increase of group-beneficial behaviors within a group. Thus, groups that possess institutions

684   can outcompete other groups in direct conflicts, and replace them, leading to increases in group-

685   level efficiency. This argument is quite plausible for early human societies for which it was

686   proposed, but perhaps not so for the majority of animal interactions that do not take place in

687   well-defined groups. However, natural selection can still align the interests of the individual with

688   that of the group in the absence of clear group structure. In particular, individuals' behavioral

689   responses to each other can provide immediate incentives for cooperation. Recently, Akçay and

690   Van Cleve (in review) have provided the general conditions for when behavioral responses in

691   structure populations can evolve to maintain group-optimal behaviors. In particular, when

692   individuals behavioral responses completely match each other's actions, the interests of

693   individuals align completely with each other, and with the aggregate group payoff. Akçay and

694   Van Cleve show that such completely matching behavioral responses can emerge through the

695   evolution of prosocial preferences in public goods games, without the need for group selection.

696   Once these prosocial preferences evolve, they render group-optimal behavioral outcomes

697   evolutionarily stable because they motivate individuals to match each other's contributions. This

698   provides the simplest possible form of institution to maintain efficient outcomes, consisting

699   simply of the evolved norm of prosocial preferences, and relying on individual's immediate

700   responses to each other. However, even in situations where immediate responses are not possible

701   (due to separation of acts in space or time), more complex elements of social structure (such as

702   the equilibrium of the law merchant described above) can still arise through selection acting on

703   individuals, and have the effect of sustaining group-level efficiency. The application of the law

704   merchant model to cleaner fish that is discussed above is an example of this: each party acts in

705   their own self-interest given what the others do, yet the arrangement of interactions solves the

706   informational and commitment problems and maintain highest mutual benefit. In short,

707   biological "institutions" might allow group-level optimization without group-level selection.

708   ## Conclusion

709   The main thesis of this review is that the theory of political and economical institutions

710   can be profitably utilized to study many phenomena in animal behavior. We did not attempt to

711   give a complete overview of institutional theory, which deals with diverse questions that range

712   from the structure and functioning of the legislative process (Baron and Ferejohn, 1989), to

713   economic development (North, 1991) and management of common resource pools (Ostrom,

714   1991). Rather, we highlighted some major areas to illustrate the approaches taken in the field

715   and how they could be applied to biology. What distinguishes institutional theory from the rest of

716   game theoretic inquiry on economics and social science is not a unique central theorem, but

717   rather an approach that seeks to understand systems of interaction as designed or evolved

718   solutions to underlying problems of exchange, production, or allocation. To use such an

719   approach in biology, we need to shift our perspective from current focus on considering each

720   individual as choosing a strategy alone in a pre-determined game, to considering the organization

721   of a social system that emerges from the interactions between individuals. From this perspective,

722   the natural history of animal interactions can be seen as similar to institutions, encoding the

723 timing and manner of how individuals interact with each other, and potentially functioning to

724 facilitate the evolution of mutual benefit through selection acting on individuals.

725 **Acknowledgements**

References

736 Akçay, E. (in review) Incentives in the family II: Behavioural dynamics and the evolution of

737 non-costly signaling.

738 Akçay, E. & Roughgarden, J. (2007) Negotiation of Mutualism: Rhizobia and legumes.

739 *Proceedings of the Royal Society B,* 274**,** 25 -- 32.

740 Akçay, E. & Roughgarden, J. (2009) The perfect family: biparental care in animals. *PLoS ONE,*

741 4**,** e7345.

742 Akçay, E. & Roughgarden, J. (2011) The evolution of payoff matrices: providing incentives to

743 cooperate. *Proceedings of the Royal Society B,* 278**,** 2198-2206.

744   Akçay, E. & Van Cleve, J. (in review) Behavioral responses in structured groups pave the way to

745   group adaptation.

746   Akçay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. (2009) The evolution of other-

747   regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of*

748   *Sciences,* 106**,** 19061-19066.

749   Austen-Smith, D. & Banks, J. S. (1996) Information aggregation, rationality, and the Condorcet

750   Jury Theorem. *American Political Science Review,* 90**,** 34-45.

751   Austen-Smith, D. & Feddersen, T. J. (2009) Information aggregation and communication in

752   committees. *Philosophical Transactions of the Royal Society B,* 364**,** 763-769.

753   Axelrod, R. & Hamilton, W. D. (1981) The evolution of cooperation. *Science,* 211**,** 1390-1396.

754   Baron, D. P. & Ferejohn, J. A. (1989) Bargaining in Legislatures. *The American Political*

755   *Science Review,* 83**,** 1181-1206.

756   Bowles, S., Choi, J.-K. & Hopfensitz, A. (2003) The co-evolution of individual behaviors and

757   social institutions. *Journal of Theoretical Biology,* 223**,** 135-147.

758   Bowles, S. & Hammerstein, P. (2003) Does Market Theory Apply to Biology? IN

759   HAMMERSTEIN, P. (Ed.) *Genetic and cultural evolution of cooperation.* Cambridge, MA, MIT

760   Press.

761   Bshary, R. (2002) Biting cleaner fish use altruism to deceive image-scoring client reef fish. *Proc*

762   *R Soc Lond B,* 269**,** 2087-2093.

763   Bshary, R. & Grutter, A. S. (2002) Asymmetric cheating opportunities and partner control in a

764   cleaner fish mutualism. *Animal Behaviour***,** 547-555.

765   Bshary, R. & Grutter, A. S. (2005) Punishment and partner switching cause cooperative

766   behaviour in a cleaning mutualism. *Biology Letters,* 1**,** 396-399.

767 Bshary, R. & Grutter, A. S. (2006) Image scoring and cooperation in a cleaner fish mutualism.

768 *Nature,* 441**,** 975-978.

769 Bshary, R. & Schäffer, D. (2002) Choosy reef fish select cleaner fish that provide high-quality

770 service. *Animal Behaviour,* 63**,** 557-564.

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

772 Clutton-Brock, T. H. (1989) Review lecture: mammalian mating systems. *Proc R Soc Lond B,*

773 236**,** 339-372.

774 Clutton-Brock, T. H. (1991) *The Evolution of Parental Care,* Princeton, NJ, Princeton University

775 Press.

776 Coase, R. (1937) The Nature of the Firm. *Economica,* 4**,** 386-405.

777 Conradt, L. & List, C. (2009) Introduction. Group decisions in humans and animals: a survey.

778 *Philosophical Transactions of the Royal Society B***,** doi:10.1098/rstb.2008.0276.

779 Conradt, L. & Roper, T. J. (2005) Consensus decision making in animals. *Trends in Ecology and*

780 *Evolution,* 20**,** 449 -- 456.

781 Conradt, L. & Roper, T. J. (2009) Conflicts of interest and the evolution of decision sharing.

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

783 Dekel, E., Ely, J. C. & Yılankaya, O. (2007) Evolution of preferences. *Review of Economic*

784 *Studies,* 74**,** 685-704.

785 Emlen, S. T. & Oring, L. W. (1977) Ecology, sexual selection and the evolution of mating

786 systems. *Science,* 197**,** 215-223.

787 Enquist, M. (1985) Communication during aggresive interactions with particular reference to

788 variation in choice of behaviour. *Animal Behaviour,* 33**,** 1152--1161.

789   Farrell, J. & Saloner, G. (1988) Coordination through committees and markets. *RAND Journal of*

790   *Economics,* 19**,** 235-252.

791   Farrell, J. & Simcoe, T. (2009) Choosing the rules for consensus standardization. *mimeo,*

792   *University of California, Berkeley*.

793   Fearon, J. D. (1995) Rationalist explanations for war. *International Organization,* 49**,** 379-414.

794   Feddersen, T. J. & Pesendorfer, W. (1997) Voting Behavior and Information Aggregation in

795   Elections With Private Information. *Econometrica,* 65**,** 1029-1058.

796   Fey, M. & Ramsay, K. W. (2009) Mechanism design goes to war: peaceful outcomes with

797   interdependent and correlated types. *Review of Economic Design,* 3**,** 233-250.

798   Fudenberg, D., Kreps, D. M. & Maskin, E. S. (1990) Repeated games with long-run and short-

799   run players. *Review of Economic Studies,* 57**,** 555-573.

800   Fudenberg, D. & Levine, D. K. (1992) Maintaining a reputation when strategies are imperfectly

801   observed. *Review of Economic Studies,* 59**,** 561-579.

802   Gardner, A. & Grafen, A. (2009) Capturing the superorganism: a formal theory of group

803   adaptation. *J Evol Biol,* 22**,** 659-71.

804   Gintis, H. (2003) The hitchhiker's guide to altruism: Gene-culture coevolution, and the

805   internalization of norms. *Journal of Theoretical Biology,* 220**,** 407-418.

806   Gintis, H. (2007) A framework for the unification of behavioral sciences. *Behavioral and Brain*

807   *Sciences,* 30**,** 1 - 61.

808   Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003) Explaining altruistic behavior in humans.

809   *Evolution and Human Behavior,* 24**,** 153 -- 172.

810   Godfray, H. C. J. (1991) Signaling of need by offspring to parents. *Nature,* 352**,** 328-330.

811   Godfray, H. C. J. & Johnstone, R. A. (2000) Begging and bleating: the evolution of parent-

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

813   Grafen, A. (1990) Biological signals as handicaps. *J. Theor. Biol.,* 144**,** 517 -- 546.

814   Greif, A. (2006) *Institutions and the Path to the Modern Economy: Lessons from Medieval*

815   *Trade,* Cambridge, Cambridge University Press.

816   Greif, A., Milgrom, P. & Weingast, B. R. (1994) Coordination, committment and enforcement:

817   the case of the merchant guild. *The Journal of Political Economy,* 102**,** 745-776.

818   Grodzinski, U., Erev, I. & Lotem, A. (2007) Can hungry nestlings be trained to reduce their

819   begging? *Behavioral Ecology,* 19**,** 116-125.

820   Grossman, S. J. & Hart, O. D. (1986) The costs and benefits of ownership: A theory of vertical

821   and lateral integration. *The Journal of Political Economy,* 94**,** 691-719.

822   Groves, T. (1973) Incentives in teams. *Econometrica,* 41**,** 617-631.

823   Groves, T. & Ledyard, J. (1977) Optimal allocation of public goods: a solution to the "free-rider"

824   problem. *Econometrica,* 45**,** 783-809.

825   Güth, W. (1995) An evolutionary approach to explaining cooperative behavior by reciprocal

826   incentives. *Int J Game Theory,* 24**,** 323-344.

827   Güth, W. & Kliemt, H. (2000) Evolutionarily stable co-operative commitments. *Theory and*

828   *Decision,* 49**,** 197-221.

829   Hauert, C. & Doebeli, M. (2004) Spatial structure often inhibits the evolution of cooperation in

830   the snowdrift game. *Nature,* 428**,** 643-646.

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

832   1078.

833 Johnstone, R. A. (1999) Signaling of need, sibling competition, and the cost of honesty.

834 *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 12644-

835 12649.

836 Johnstone, R. A. & Grafen, A. (1992) The continuous Sir Phillips Sidney game: A simple model

837 of biological signaling. *Journal of Theoretical Biology,* 156**,** 215-234.

838 Klein, B., Crawford, R. G. & Alchian, A. A. (1978) Vertical interaction, appropriable rents and

839 the competitive contracting process. *Journal of Law and Economics,* 21**,** 297-326.

840 Klemperer, P. (1999) Auction theory: a guide to the literature. *Journal of Economic Surveys,* 13**,**

841 227-286.

842 Kreps, D. M. & Wilson, R. (1982) Reputation and imperfect information. *Journal of Economic*

843 *Theory,* 27**,** 253-279.

844 Lehmann, L. & Keller, L. (2006) The evolution of altruism and cooperation -- a general

845 framework and a classification of models. *Journal of Evolutionary Biology,* 19**,** 1365 -- 1376.

846 List, C. (2004) Democracy in animal groups: a political science perspective. *Trends in Ecology*

847 *& Evolution,* 19**,** 168-169.

848 Maynard Smith, J. & Harper, D. G. C. (1988) The evolution of aggression: can selection

849 generate variability? *Philosophical Transactions of the Royal Society B,* 319**,** 557-570.

850 Maynard Smith, J. & Parker, G. A. (1976) The logic of asymmetric contest. *Animal Behaviour,*

851 24**,** 159-175.

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

853 Mcfarland, D. & Houston, A. (1981) *Quantitative ethology: the state space approach,* Boston,

854 MA, Pitman Advanced Pub. Program.

855 Mcnamara, J. M., Gasson, C. E. & Houston, A. I. (1999) Incorporating rules for responding into

856 evolutionary games. *Nature,* 401**,** 368-371.

857 Meirowitz, A. (2006) Designing Institutions to Aggregate Preferences and Information.

858 *Quarterly Journal of Political Science,* 1**,** 373-392.

859 Meirowitz, A. & Sartori, A. (2008) Strategic uncertainty as a cause of war. *Quarterly Journal of*

860 *Political Science,* 3**,** 327-352.

861 Milgrom, P. M. & Weber, R. J. (1982) A theory of auctions and competitive bidding.

862 *Econometrica,* 50**,** 1089-1122.

863 Milgrom, P. R., North, D. C. & Weingast, B. R. (1990) The role of institutions in the revival of

864 trade: the law merchant, private judges and the champagne fairs. *Economics and Politics,* 2**,** 1-

865 23.

866 Myerson, R. B. (1979) Incentive-compatibility and the bargaining problem. *Econometrica,* 47**,**

867 61-73.

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

869 Myerson, R. B. & Satterthwaite, M. A. (1983) Efficient mechanisms for bilateral trading.

870 *Journal of Economic Theory,* 29**,** 265-281.

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

872 Noë, R. & Hammerstein, P. (1995) Biological markets. *Trends in Ecology & Evolution,* 10**,** 336 -

873 - 339.

874 Noë, R., Van Hooff, J. A. R. A. M. & Hammerstein, P. (Eds.) (2001) *Economics in nature:*

875 *social dilemmas, mate choice and biological markets,* Cambridge, Cambridge University Press.

876 Nöldeke, G. & Samuelson, L. (1999) How costly is the honest signaling of need? *Journal of*

877 *Theoretical Biology,* 197**,** 527–539.

878 North, D. C. (1991) *Institutions, institutional change and economic performance,* Cambridge,

879 Cambridge University Press.

880 Ok, E. A. & Vega-Redondo, F. (2001) On the Evolution of Individualistic Preferences: An

881 Incomplete Information Scenario. *Journal of Economic Theory,* 97**,** 231-254.

882 Okasha, S. (2006) *Evolution and the Levels of Selection.,* Oxford, Oxford University Press.

883 Orians, G. H. (1969) On the evolution of mating systems in birds and mammals. *The American*

884 *Naturalist,* 103**,** 589-603.

885 Ostrom, E. (1991) *Governing the commons: the evolution of institutions for collective action,*

886 Cambridge, Cambridge University Press.

887 Pereira, H., Bergman, A. & Roughgarden, J. (2003) Socially stable territories: The negotiation of

888 space by interacting foragers. *The American Naturalist,* 161**,** 143-152.

889 Riley, J. G. & Samuelson, W. F. (1981) Optimal Auctions. *The American Economic Review,* 71**,**

890 381-392.

891 Rodríguez-Gironés, M. A., Cotton, P. A. & Kacelnik, A. (1996) The evolution of begging:

892 signalling and sibling competition. *Proceedings of the National Academy of Sciences,* 93**,** 14637-

893 146641.

894 Rodríguez-Gironés, M. A., Zúñiga, J. M. & Redondo, T. (2002) Feeding experience and relative

895 size modify the begging strategies of nestlings. *Behavioral Ecology,* 13**,** 782-785.

896 Roughgarden, J. (2009) *The Genial Gene: Deconstructing Darwinian Selfishness*, Berkeley:

897 University of California Press.

898 Roughgarden, J., Oishi, M. & Akçay, E. (2006) Reproductive social behavior: Cooperative

899 games to replace sexual selection. *Science,* 311**,** 965-970.

900	Roughgarden, J. & Song, Z. (in press) Incentives in the Family I: The Family Firm, An

901	Evolutionary/Economic Theory for Parent-Offspring Relations. IN NARVAEZ, D., PANKSEPP,

902	J., SCHORE, A. & GLEASON, T. (Eds.) *Human Nature, Early Experience and the Environment*

903	*of Evolutionary Adaptedness.* New York, Oxford University Press.

904	Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. (2004) The evolution of cooperation.

905	*Quarterly Review of Biology,* 79**,** 135-160.

906	Stamps, J. A. & Krishnan, V. V. (2001) How territorial animals compete for division of space: a

907	learning-based model with unequal competitors. *The American Naturalist,* 157**,** 154-169.

908	Taylor, P. D. & Day, T. (2004) Stability in negotiation games and the emergence of cooperation.

909	*Proc. R. Soc. B,* 271**,** 669-674.

910	Vickrey, W. (1961) Counterspeculation, auctions and compatitive sealed tenders. *The Journal of*

911	*Finance,* 16**,** 8-37.

912	Weibull, J. & Salomonsson, M. (2006) Natural selection and social preferences. *Journal of*

913	*Theoretical Biology,* 239**,** 79-92.

914	West, S. A., Griffin, A. S. & Gardner, A. (2007) Evolutionary explanations for cooperation.

915	*Current Biology,* 17**,** R661 -- R672.

916	Williamson, O. E. (1979) Transaction-cost economics: the governance of contractual relations.

917	*Journal of Law and Economics,* 22**,** 233-261.

918	Wolff, J. O. & Macdonald, D. W. (2004) Promiscuous females protect their offspring. *Trends in*

919	*Ecology & Evolution,* 19**,** 127 -- 134.

920	Worden, L. & Levin, S. A. (2007) Evolutionary escape from the prisoner's dilemma. *Journal of*

921	*Theoretical Biology,* 245**,** 411 -- 422.

922   Young, H. P. (2008) Social norms. IN DURLAUF, S. N. & BLUME, L. E. (Eds.) *The New*

923   *Palgrave Dictionary of Economics, Second edition.* London, Macmillan.

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

925   53**,** 205-214.

926

927