vol. 168, no. 6 the american naturalist december 2006

E-Article

Why Animals Lie: How Dishonesty and Belief Can Coexist in a Signaling System

Jonathan T. Rowell,[1,*] Stephen P. Ellner,[2,†] and H. Kern Reeve[3,‡]

1. Center for Applied Mathematics, Cornell University, Ithaca, New York 14853;
2. Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853;
3. Department of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853

Submitted February 24, 2006; Accepted June 26, 2006; Electronically published November 2, 2006

abstract: We develop and apply a simple model for animal communication in which signalers can use a nontrivial frequency of deception without causing listeners to completely lose belief. This common feature of animal communication has been difficult to explain as a stable adaptive outcome of the options and payoffs intrinsic to signaling interactions. Our theory is based on two realistic assumptions. (1) Signals are "overheard" by several listeners or listener types with different payoffs. The signaler may then benefit from using incomplete honesty to elicit different responses from different listener types, such as attracting potential mates while simultaneously deterring competitors. (2) Signaler and listener strategies change dynamically in response to current payoffs for different behaviors. The dynamic equations can be interpreted as describing learning and behavior change by individuals or evolution across generations. We explain how our dynamic model differs from other solution concepts from classical and evolutionary game theory and how it relates to general models for frequency-dependent phenotype dynamics. We illustrate the theory with several applications where deceptive signaling occurs readily in our framework, including bluffing competitors for potential mates or territories. We suggest future theoretical directions to make the models more general and propose some possible experimental tests.

Keywords: signaling, communication, deception, eavesdropping, evolutionary game theory, strategy dynamics.

* Present address: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, Tennessee 37996; e-mail: jtrowell@ec.rr.com.

† Corresponding author; e-mail: spe2@cornell.edu.

‡ E-mail: hkr1@cornell.edu.

Am. Nat. 2006. Vol. 168, pp. E180–E204.

Early studies of animal signaling frequently presumed that signals to conspecifics evolved to facilitate and coordinate social interactions by "the cooperative exchange of reliable information" (Searcy and Nowicki 2005, p. 2; review in Johnstone 1998). Game theory revolutionized the understanding of signals by explaining communication as a strategic interaction (Maynard Smith and Price 1973; Maynard Smith 1974, 1979; Maynard Smith and Parker 1976). The game-theoretic view demonstrated that if a signaler could benefit by deceiving a conspecific, then the communication system was susceptible to invasion by deceptive mutants. Then listeners, faced with increasingly deceptive signals, would evolve to disregard the signal (Johnstone 1998). In the end everyone would be lying and no one listening.

This is a problematic conclusion because honest communication is well documented (Johnstone 1998). Johnstone (1998, p. 95) succinctly states the problem as follows: "How is informative communication (the transfer of reliable information between two or more individuals) possible when signalers stand to gain by deceit?"

The first explanation offered was that signals of individual quality are inherently reliable because they are functionally tied to the advertised qualities (Maynard Smith and Parker 1976; Zahavi 1977*b*; Maynard Smith 1979; Wiley 1983; Maynard Smith and Harper 1988; Hughes 2000). Alternatively, a state in which honesty is dominant may be a "temporary phase in [an] arms race" (Johnstone 1998, p. 96). Under this interpretation, signal systems are constantly being created and then destroyed by selection.

Zahavi's (1975, 1977*a*, 1981, 1987) handicap principle held that dishonest signals of high quality were discouraged because a low-quality signaler either paid a higher cost than did a high-quality signaler or would be unable to take full advantage of the benefit that a high-quality signaler would receive. Zahavi's hypothesis encountered limited acceptance until a mathematically sound foundation was produced by Grafen (1990). Following this, others showed how the handicap principle could be applied to quality advertisement (Grafen 1990; Johnstone 1994, 1995), need advertisement (Godfray 1991, 1995; Maynard

Smith 1991, 1994; Johnstone and Grafen 1992), or both (Adams and Mesterton-Gibbons 1995; Godfray 1995). The accepted interpretation of animal communication has therefore gone through a cycle: implicit honesty → no honesty and no long-term communication → (and now) cost-reinforced honesty.

However, dishonesty is also widespread. For example, the meral spread displays of stomatopods (*Gonodactylus bredini*) are aggressive signals used in territory defense and are given by both newly molted and intermolt individuals, even though newly molted individuals cannot defend themselves adequately (Steger and Caldwell 1983; Adams and Caldwell 1990). In green tree frogs *Rana clamitans* some small males exaggerate their quality by lowering their acoustic pitch to resemble that of larger males (Bee et al. 2000). False alarm signals may be given to divert rivals from food sources or mating opportunities, as in the shrikes *Lanio versicolor* and *Thamnomanes schistogynus* (Munn 1986). Deception has been observed in all primate groups, and differences in deception rate among primate species correlate with neocortex size, suggesting that benefits from deception may have been a driver of neocortex expansion (Byrne and Corp 2004).

The handicap principle does not completely preclude deception (Grafen 1990; Johnstone and Grafen 1993; Adams and Mesterton-Gibbons 1995; Searcy and Nowicki 2005). If most signals are honest, if the cost to deceptive signalers is low when they are rare (Gardner and Morris 1989), or if the cost of ignoring an honest signal is high (Searcy and Nowicki 2005), occasional deceit can be stable. However, there is, at present, no widely accepted general explanation for how substantial levels of dishonesty can persist without destabilizing communication.

Several hypotheses have been advanced in which the frequency of deception is determined by factors extrinsic to the signaling interaction (Johnstone 1998). Similar to the handicap principle is the notion that listeners must pay an "assessment cost" in order to determine a signal's honesty (Dawkins and Guilford 1991; Getty 1997). For example, in male-male combat, an opponent's true strength is revealed only in actual fighting. Other ideas are that deception occurs as a result of imperfect relationships between signal structure and competitive ability (Dawkins and Guilford 1991) or high benefits to some signalers when deceptive displays are successful (Adams and Mesterton-Gibbons 1995; Hughes 2000). Hughes (2000) argued that dishonest signalers can use real-world variability and noise to mask themselves (see also Johnstone and Grafen 1993; Adams and Mesterton-Gibbons 1995; Godfray 1995).

A second type of hypothesis is that a mix of deception and honesty can result from differences between signalers. Johnstone and Grafen (1993) presented a model with two types of signaler differing in their payoffs such that at the evolutionarily stable strategy (ESS) one type was honest and the other deceptive. The frequency of deception in this model is determined largely by the frequency of different signaler types in the population. Similarly, Kokko (1997) developed models for mating signals in which the ESS could involve honesty and deception at different ages. Heterogeneity among signalers also underlies the well-known models for meral spread displays in stomatopods. At the ESS the strongest and the weakest individuals both give threat displays, which are honest for the former but deceptive for the latter (Adams and Mesterton-Gibbons 1995).

In this article we present a new hypothesis for the persistence of partially deceptive communication. In our model the occurrence and frequency of deception derive from intrinsic properties of the signaling interaction rather than from extrinsic factors such as noise and are not the result of honest and deceptive signals being employed by distinct sets of signalers with different cost-benefit relationships. However, we are not intending any criticism of previous hypotheses, nor do we raise any conflicts with previous theoretical work.

The first part of our hypothesis is that signal recipients (rather than signalers) are heterogeneous. The evolution of signals is conventionally considered to be driven by pairwise interactions between a signaler and a recipient (Doutrelant et al. 2001). However, there is growing evidence (reviewed by Searcy and Nowicki [2005], chap. 5) that signals are often received by individuals besides the primary receiver (Wiley 1983; McGregor 1993; McGregor and Peake 2000; Doutrelant et al. 2001). By way of third-party signal interception and observation of interactions, individuals can gain information about others with whom they are not directly communicating or interacting (McGregor and Dabelsteen 1996; Naguib and Todt 1997; Oliveira et al. 1998; Otter et al. 1999; Doutrelant and McGregor 2000; Earley and Dugatkin 2002). We develop a model in which different listeners or listener types can respond to a given signal in different ways.

The second part of our hypothesis is that strategies are dynamic, continuously changing in response to the current behavior of others. We begin with a classical game theory model but then use the payoff matrix to define a model for the dynamics of signaling strategies rather than seeking conventional ESS solutions for the matrix game with the specified payoffs. Our approach also differs from conditional ESSs for repeated games (such as tit-for-tat in iterated Prisoners' Dilemma) that allow individual behavior to vary over time depending on its circumstances. In our approach it is the strategies themselves—the sets of rules that govern each decision about how to signal or respond—that are changing over time in response to current conditions and not just the behaviors. Our model can be

interpreted as operating on either evolutionary (across-generation) or behavioral (within-generation) timescales and is mathematically similar to the replicator and imitation models for behavior dynamics (Hofbauer and Sigmund 1998, 2003).

The article is organized as follows. In "A Dynamic Model for Signaling Games" we derive a minimal model for signaling with two listener types and discuss the model's limitations. We then describe some general properties of the model ("Reduced Models: Static Strategies and One-Listener Games") and present some specific applications to animal signaling interactions ("Applications to Animal Signaling"). Our models reveal that for a signaler facing a mixed population of listeners, some aligned with and some opposed to the signaler's interests, a mixed signaling strategy may be best. For example, partial deception may allow one signal to elicit different responses from different listeners: for some the signal is reliable enough that they gain most by treating it as honest, while for others with different payoffs the signal is too unreliable to be trusted. The combination of a third-party listener and dynamic strategy updating readily predicts some degree of partial deception and some previously unsuspected dynamical outcomes. In "Discussion" we summarize our results and suggest some possible experimental tests and directions for future research to improve the model's generality and realism. To make the main text accessible, technical derivations and proofs are placed in a series of appendixes.

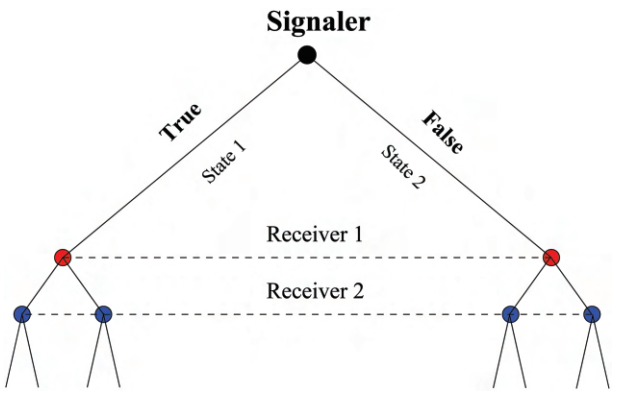

Figure 1: Tree diagram of the minimal two-listener signaling game. Each node (indicated by a colored circle) represents a decision moment for one of the players. The signaler, aware of the true state of nature, produces a signal that may be either true or false (*black node*), depending on the state of nature. Then receivers 1 (*red nodes*) and 2 (*blue nodes*) simultaneously decide on courses of action. Nodes connected by a dashed line represent an information set—the player in question knows that he is at one of the nodes in the set but does not know which specific one. In this case the receivers know that a signal was produced but do not know whether the signal was true or false and do not know which action was chosen by the other receiver. Payoffs occur at the open terminal branches of the tree.

A Dynamic Model for Signaling Games

For this initial study of two-listener signaling games, we have aimed for simplicity rather than generality. We consider a minimal model in which signaling is a game between three players or types of player, each having a binary choice of action (fig. 1). The signaler may either truthfully give a signal that describes the actual environment and/or condition of the signaler or send the same signal "dishonestly." So when a false signal is transmitted, the environment or signaler condition is different from when the signal is truthful. Each listener can react as if the signal were true or else react under the assumption that it is false. The actions corresponding to belief and disbelief need not be identical for the two listeners. The two listeners are assumed to act "simultaneously" in the game-theoretic sense, meaning that each must choose its course of action without knowing the other's action. The signaler's level of truthfulness t is a variable in the interval $[0, 1]$ and represents the fraction of signals that are truthful. Likewise, the variables p and q represent the fraction of signals believed by listeners 1 and 2, respectively. Our model can also be interpreted as applying to populations of individuals, each of whom plays a pure strategy. Under this interpretation t is the fraction of signalers giving honest signals, and p and q are the fractions of recipients of each type who believe the signal to be honest.

Our signaling game is analogous to Arnold's (1978) model for Batesian mimicry, in which two possible situations are indistinguishable to a predator (exploitable vs. harmful prey). In our model, signal receivers are faced with two indistinguishable states of nature corresponding to the signal being truthful or dishonest.

Our model ignores payoffs or information from occasions when signals are not produced. This is not always legitimate. Consider predator alarm signals. A signaler's strategy has two components—the behaviors when predators are absent and when they are present—and listener behavior is likely to be influenced by the overall frequency of predator attack, not just the frequency of attack when an alarm has been given. However, in this article we consider only situations fitting figure 1 in the sense that "action" occurs only in response to signals. More general game structures will be considered elsewhere (J. T. Rowell, unpublished manuscript). Our payoff functions assume a constant payoff for each possible type of interaction and pair- or tripletwise interactions among groups drawn at random from the pools of each player type. These as-

sumptions lead to simple payoff functions reminiscent of classical matrix games such as hawk-dove.

The Signaling Game

The two independent choices for each of the three players produce a total of eight possible payoff scenarios. The payoff structure can be represented by the following normal form payoff matrix where A_i represents the payoff for listener 1 (L1), B_i is the payoff for listener 2 (L2), and C_i is the payoff for the signaler.

	Truthful Signal, $t = 1$	
L1 Believe, $p = 1$	A_1, B_1, C_1	A_2, B_2, C_2
L1 Not Believe, $p = 0$	A_3, B_3, C_3	A_4, B_4, C_4
	L2 Believe, $q = 1$	L2 Not Believe, $q = 0$

(1)

	Untruthful Signal, $t = 0$	
L1 Believe, $p = 1$	A_5, B_5, C_5	A_6, B_6, C_6
L1 Not Believe, $p = 0$	A_7, B_7, C_7	A_8, B_8, C_8
	L2 Believe, $q = 1$	L2 Not Believe, $q = 0$

The payoff to mixed strategies is computed as the average payoff over all possible combinations, weighted by their expected frequency, with interactants drawn randomly from the population (e.g., the frequency of truthful signals that are disbelieved by both listeners is $t(1 - p)(1 - q)$).

General Dynamic Model

We now deviate from classical game theory by using the payoffs to create a dynamical system modeling behavioral evolution or learning. This dynamic model allows us to study the potential for transient behavior in signaling systems.

Each player has a binary choice of actions, which can be associated with the values 0 and 1 (a mixed strategy then has a value $x \in [0, 1]$ representing the probability of choosing action 1). Let the expected average payoff rates for the three types of players be denoted $f_i(\mathbf{x})$, where $\mathbf{x} = (x_1, x_2, x_3) = (p, q, t)$. The general version of our model is

$$x_i' = \frac{\partial f_i}{\partial x_i} x_i(1 - x_i), \tag{2}$$

where a prime denotes the derivative with respect to time.

We have stated the model before its assumptions because several different assumptions lead to the same trait-level dynamics, albeit with different interpretations. First, we can interpret equation (2) as describing evolutionary change in a trait under selection. The simplest such model is to posit that the behavior alternatives (truth vs. deception, belief vs. doubt) are Mendelian traits corresponding to two alleles at a locus. Our model is then obtained as the classical continuous-time model for allele frequency dynamics, $p_i' = p_i(1 - p_i)(\partial \overline{W}/\partial p_i)$, where $\overline{W}$ is the mean fitness. (Note, however, that under frequency dependence the fitness gradient $\partial \overline{W}/\partial p_i$ has to be interpreted as the change in mean fitness for a small subpopulation with a trait mutation in a large population that remains unchanged. For our model this is the mean fitness for a small subpopulation with an altered frequency of different behaviors, but the payoff for each behavior is still determined by the behavior frequencies in the general population.)

Also, three different approaches to modeling quantitative trait evolution lead to fitness-gradient trait dynamics as in equation (2), and "all support the use of such dynamics as a rough approach to understanding evolutionary questions involving frequency dependence" (Abrams 2005, p. 1165). These approaches are quantitative genetics theory (QG), adaptive dynamics (AD), and the *G*-function approach of Brown and Vincent (Abrams 2005). All of these can produce model (2), under suitable assumptions. The QG approach is the generalization to frequency-dependent selection of the classical "breeder's equation" ($R = h^2S$); it is most appropriate for multilocus traits in sexually reproducing populations. In the QG model the fitness gradient is the partial derivative of individual fitness with respect to individual trait value (evaluated at the population trait mean), and the selection response is the product of the fitness gradient and the additive genetic variance (Abrams et al. 1993; if several traits are simultaneously evolving in a population, the response of each trait is modified by genetic correlations, which we assume to be absent—this is not necessarily true but is possible because the traits are functionally unrelated). To reach equation (2) in the QG approach, we regard the state variables (p, q, t) as trait means with possible trait values 0 and 1. The logistic terms in equation (2) are then the trait variances, which (assuming constant heritability) are proportional to the additive genetic variance. We then scale time so that the constant of proportionality is 1, making the assumption that the same scaling works for all player types. Under the same assumptions the *G*-function also produces equation (2), as an instance of equation (3) of Cohen et al. (1999). We can also assume instead that individuals play mixed strategies, with (p, q, t) as the frequencies of choosing behavior 1. The QG and *G*-function models can still apply if we assume that the additive genetic variance is proportional to $x_i(1 - x_i)$, to model the fact that as the population mean approaches either 0 or 1, the variance must drop to 0 (as done, e.g., by Abrams [1999] with a different functional form). Similarly, in the so-called

canonical equation of AD the fitness gradient is multiplied by the variance of mutation effects, and one might rationalize $x_i(1 - x_i)$ as a model for mutation variance by assuming that as the trait approaches its minimum or maximum values, there are fewer and fewer options for change. None of these stories will ever hold exactly, but their convergence on one outcome suggests that equation (2) has some robustness as a rough approach.

Our model also has a second interpretation in which signaling strategies change over time due to learning. As Abrams (2005, p. 1165) notes, "there is a less rigorous but still persuasive case that many behavioral traits can also be roughly described by the same general family of [fitness gradient] models." In this interpretation, our equations have the structure of a learning or imitation model (Hofbauer and Sigmund 1998, 2003). The logistic term $x_i(1 - x_i)$ represents the frequency of random encounters between individuals playing different strategies (with time scaled so that encounters occur at a rate of one per unit time, and the encounter rate is assumed to be the same for all types of player). Encounters with individuals playing different strategies allow individuals to compare payoffs. If an individual observes that a different behavior has a higher average payoff, it can either switch to that strategy (if we imagine populations composed of a mixture of pure strategists) or increase the frequency with which it uses the more rewarding option (if we imagine populations of mixed strategists). In either case, we assume that the higher payoff option becomes more frequently used, at a rate proportional to the fitness benefit for a switch to that option. Models of this kind with a single type of player, rather than three distinct roles, have been studied extensively (Hofbauer and Sigmund 2003, sec. 3.2).

Model (2) is conceptually similar to the classical replicator dynamics model for evolutionary games (Hofbauer and Sigmund 1998; Nowak and Sigmund 2004). But in several of the situations that we would like our model to cover, such as bluffing in signals to potential mates or competitors, both signalers and listeners are potentially playing a mixed strategy. The classical replicator model assumes that the population is a mix of individuals playing pure strategies, and deductions about corresponding mixed-strategy ESSs hold only at the model's steady states. We therefore prefer to base our model on the learning dynamics equations so that (as under the evolutionary dynamics interpretation) the same model can be applied to either pure or mixed strategies.

Signaling Game Dynamics

In our specific models for signaling games, the payoff function for each player type is a linear function of its own strategy variable; that is, each f_j is linear in x_j. The fitness gradients in equation (2) can therefore be computed as the expected payoff difference between the two pure strategies 1 and 0, given the strategies of the other players. As an example, the matrix Δt of payoff differences for the signaler would be

$$\Delta t = \begin{array}{c|cc} & q = 1 & q = 0 \\ p = 1 & C_1 - C_5 & C_2 - C_6 \\ p = 0 & C_3 - C_7 & C_4 - C_8 \end{array} = \begin{array}{|c|c|} \hline \Delta C_1 & \Delta C_2 \\ \hline \Delta C_3 & \Delta C_4 \\ \hline \end{array} \qquad (3)$$

The payoffs in this matrix can be used to compute the expected payoff difference for truthfulness versus dishonesty, given that the listeners are using the mixed strategies p and q:

$$\begin{aligned} r_t(p, q) &= E(\Delta t | p, q) \\ &= \Delta C_1 pq + \Delta C_2 p(1 - q) \\ &\quad + \Delta C_3 q(1 - p) + \Delta C_4(1 - p)(1 - q). \end{aligned} \qquad (4)$$

Using this payoff differential in our general model (eq. [2]), we obtain the dynamic equation for signaler truthfulness t:

$$\begin{aligned} t' = [&\Delta C_1 pq + \Delta C_2 p(1 - q) + \Delta C_3 q(1 - p) \\ &+ \Delta C_4(1 - p)(1 - q)]t(1 - t), \end{aligned} \qquad (5)$$

where a prime denotes the derivative with respect to time. We construct the differential equations describing the behavior of belief levels p and q in exactly the same way. These two equations are

$$\begin{aligned} p' = [&\Delta A_1 tq + \Delta A_2 t(1 - q) + \Delta A_5 q(1 - t) \\ &+ \Delta A_6(1 - t)(1 - q)]p(1 - p), \end{aligned} \qquad (6)$$

$$\begin{aligned} q' = [&\Delta B_1 pt + \Delta B_3 t(1 - p) + \Delta B_5 p(1 - t) \\ &+ \Delta B_7(1 - p)(1 - t)]q(1 - q), \end{aligned} \qquad (7)$$

where the coefficients are $\Delta A_i = A_i - A_{i+2}$ and $\Delta B_i = B_i - B_{i+1}$. The terms in brackets will be referred to as $r_p(t, q)$ and $r_q(t, p)$, respectively.

The overall state of the model (p, q, t) can be thought of as a point within or on the surface of the unit cube $[0, 1] \times [0, 1] \times [0, 1]$, which we will call the signaling cube. Figure 2 labels various parts of the cube for future reference. The cube has a variety of possible equilibrium structures, which are explained in appendix A. The eight vertices are always equilibria as a result of the logistic growth terms, but this is reasonable biologically. In a com-

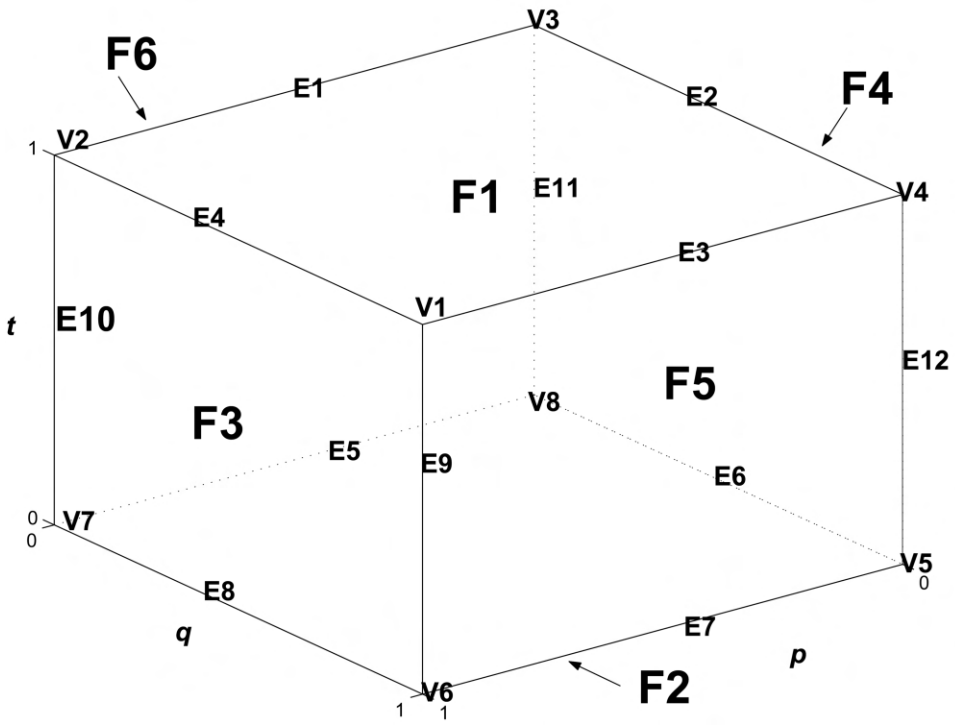

Figure 2: Regions of the signaling cube labeled for reference. Each face represents a pure strategy by one type of player (complete honesty or dishonesty by the signaler, complete belief or disbelief by one type of listener). Each edge represents a pure strategy by two types of player, and each vertex represents a pure strategy by all players. Vertex *V*1 is the ethological vertex $t = p = q = 1$, with complete honesty and belief; the opposite corner *V*8 is the breakdown vertex $t = p = q = 0$, with complete dishonesty and disbelief. Each face, edge, and vertex is an invariant set for the model because player types with pure strategies do not have any trait variance on which selection can act.

pletely homogeneous population, strategies cannot evolve without a mutation that perturbs the system away from the vertex, even if higher-fitness strategies are possible. In the learning-dynamics interpretation, vertex equilibria are situations where individuals do not see any examples of alternative behaviors and therefore do not change their own. The vertex (1, 1, 1) will be referred to as the "ethological equilibrium" because there is complete honesty and total belief. Similarly, (0, 0, 0) represents signaling breakdown because there is complete dishonesty and total disbelief.

Reduced Models: Static Strategies and One-Listener Games

Our dynamic model adds two realistic features that distinguish it from classical signaling models: the presence of two distinct listener types and a focus on continuous dynamic updating or evolution of signaling strategies. The predictions that we derive below for particular signaling interactions result from the combination of these assumptions rather than from either one alone. To justify this claim, we have to compare our model against properties of the "reduced" models that omit one of the added features.

Static Strategies

A conventional matrix game involves all participants choosing their strategies simultaneously (as noted above, this is really the assumption that players must act without knowing what actions others have chosen). A Nash equilibrium is achieved when no player can improve his or her expected payoff by independently changing his or her strategy. The equilibrium is strict if every strategic change results in an absolute decrease of payoff. This last condition is the most common form of the noninvasibility condition that characterizes an ESS. We refer to these as static solutions or strategies because they are based on individuals repeating the same strategy (pure or mixed)

over many repetitions of the game—this is what justifies the "expected payoff" criterion.

The relationship between our strategy dynamics model and these static solution concepts is analyzed in appendix B. The key results are that any Nash equilibrium (as defined above) (*a*) is a fixed point of the dynamic strategy model and (*b*) is a local ESS if and only if it is a locally stable equilibrium in the dynamic model with no eigenvalues having 0 real part. Thus, our dynamic model is an extension of the classical solution concepts for matrix games (and always includes the classical solutions as possible outcomes) rather than a completely different solution concept. Property (*a*) is proved for the general model (eq. [2]) and is not just a special property of the specific payoff functions in equations (5)–(7). Similar results for other models of behavior dynamics are reviewed by Hofbauer and Sigmund (2003).

A related concept from game theory is a repeated game, where a game is played ad infinitum but players can use strategies that depend on time or on past events. The best-known example of this type is the tit-for-tat strategy in repeated Prisoners' Dilemma. A player's behavior sequence during a repeated game can be evocative of evolutionary change or learning but is fundamentally different: although the behavior sequence is time varying (and may include conditional responses such as retaliation), each player's "rule book" is time invariant. In our approach, the strategy itself changes in response to the changing environment. Our approach also differs from classical solutions for repeated games in that there is no consideration of long-term future payoffs—strategies are updated in ways that improve immediate payoffs under current conditions.

One-Listener Games

The one-listener reduced game is analyzed in appendix C. This model has a variety of different behaviors depending on parameter values, so rather than trying to summarize those results here, we will refer to appendix C results when they are needed.

Applications to Animal Signaling

We now consider some specific signaling interactions that may be modeled as a three-player game using our model. Each is found to have solutions with mixed levels of honesty on the part of the signaler. Although we have noted several restrictive assumptions in our model, the situations that we selected to model here are ones for which we believe that our model provides a reasonable initial description. We present here two examples motivated by animal signaling. In appendix D we present a third example involving human communication, which is somewhat contrived but may be useful for the experimental tests using human subjects that we propose in "Discussion."

Many of the results below were derived by local linear stability analysis of fixed points. For the sake of brevity we omit these calculations; details are given by Rowell (2003). Also, we do not explicitly discuss the corresponding games with static strategies because the results in appendix B imply the most important conclusion: a classical ESS strategy appears in the dynamic models as convergence to a stable equilibrium, and any other kind of behavior dynamics is an additional prediction of our dynamic model.

Raiders and Ambushes

In this section we present an example of our general model that serves as a paradigm for many of its important properties. The signaler is a territory-holding male; the listeners are females whom the signaler would like to attract and satellite males who want to sneak matings with females attracted by the mating calls. For the signaler, the ideal signal would simultaneously attract females and deter satellites. Our model shows that a partially deceptive signal can be a means for achieving that objective.

Territory-holding males often use signals to attract nearby females into their territory. The mating signal may be produced concurrent with a defense signal or independently. In many species nonterritorial males can still gain access to mating opportunities. One alternative mating tactic for nonterritorial males is mate interception (e.g., Lucas et al. 1996). This behavior is seen in many taxonomic groups, but it is especially common in anurans. Once a call is produced by an established territorial bull, the satellite raider will enter the territory and attempt a forced mating with any female located. Even if the territory holder ejects the raider after a copulation occurs, the holder's mating success will be reduced through sperm competition.

The status of being a territorial male or a satellite male may be size determined or an alternative phenotypic expression, or it may be the strategic choice of a male who is capable of switching tactics. Experimental evidence from Lucas et al. (1996) suggests that raider natterjack toads *Bufo calamita* are smaller than callers. This supports the hypothesis that the satellite role is a conditional ESS (Arak 1988).

In this section we apply our general model to the interaction between territory-holding males, satellite males, and females. Specifically, we examine whether signalers might benefit from giving false mating calls for the express

purpose of ambushing satellite males. "False" means that any individual who responds to the call—male or female—is immediately attacked when he or she enters the territory. Puzzling cases of male aggression toward sexually receptive females are known to occur in a wide variety of animals, for example, rove beetles (Alcock and Forsyth 1988), seed-eating lygaeid bugs (McLain and Pratt 1999), cichlid fish (Clement et al. 2005), and Sulawesi crested black macaques (Reed et al. 1997). Such cases may reflect sexual coercion (McLain and Pratt 1999) or paternity assurance (Alcock and Forsyth 1988).

We propose here another possible explanation for these ambush attacks on potential mates: they deter raids by satellite males. Although it is counterproductive to attack and injure a female, pausing to evaluate the individual entering the territory may lead to a lost opportunity to strike a raiding satellite male with low cost and high odds of success. Unconditional ambush may be part of a successful mixed strategy if a low rate of ambushing produces a splitting of behavior between satellites and females so that satellites are deterred from entering the territory but females are not.

Payoff Matrices. We assume that caller and satellite are nonswitchable roles, with callers being larger and more likely to win in a conflict with a satellite. With satellite males as listener 1 and females as listener 2, the payoff matrix for our ambush game model is

	Caller: Attract Mate	
Satellite: Raid	$N, N, -E$	$-F, b_2, -(F\phi + E)$
Satellite: Avoid	$b_1, M, M - E$	$b_1, b_2, -E$
	Female: Enter	Female: Avoid

(8)

	Caller: Ambush	
Satellite: Raid	$-A/2, -A/2, 0$	$-A, b_2, 0$
Satellite: Avoid	$b_1, -A, 0$	$b_1, b_2, 0$
	Female: Enter	Female: Avoid

Model parameters are summarized in table 1. The rationale behind the payoffs is as follows. First consider a caller making an honest attempt to attract a mate. If a satellite raids and the female enters the territory (first row, first column of matrix), the satellite intercepts the female, obtaining a reproductive payoff N (as does the female), and the caller incurs the cost $E \geq 0$ for honest signaling. A value of $E > 0$ is not an intrinsic signal cost but a cost for getting into a state of readiness for mating (e.g., reduced immune function or increased exposure to predators). If the satellite raids and the female avoids the territory (first row, second column), then the satellite incurs the fighting cost F for encountering only the caller, the female gets its background reproductive payoff b_2, and the caller incurs both the honest signal cost and the cost of fighting with the satellite ($E + F\phi$, with $0 \leq \phi \leq 1$ because the caller is not weaker than the satellite). If the satellite avoids raiding and the female enters the territory (second row, first column), then the satellite receives its background reproductive payoff b_1, and the female and the caller both receive the payoff M from mating with each other. If the satellite avoids raiding and the female avoids the territory (second row, second column), then the satellite and female receive their background payoffs b_1 and b_2, respectively, and the caller incurs the honest signal cost E. We can assume that $N > b_1$ and $M > b_2$: a successful mating is better than the background opportunities.

Next consider that the caller attempts to ambush the satellite, that is, provides a deceptive mate-attraction signal. If the satellite raids and the female enters the territory (first row, third column), then the satellite and female each absorb expected damage $A/2$ from the ambushing caller. The ambushing caller suffers no cost of fighting, that is, the assumed benefit of preemptive ambush. If the satellite raids and the female avoids the territory (first row, fourth column), then the satellite absorbs all of the damage A from the ambushing caller, the female gets the background payoff b_2, and the caller experiences negligible payoff change. If the satellite avoids raiding and the female enters the territory (second row, third column), then the satellite receives its background reproductive payoff b_1, the female now absorbs all of the damage A from the ambushing caller, and the caller experiences negligible payoff change. Finally, if the satellite avoids raiding and the female avoids the territory (second row, fourth column), then the satellite and the female receive their background reproductive payoffs b_1 and b_2, respectively, and the caller experiences negligible payoff change.

Table 1: Parameters of the ambush game

Parameter	Description
N, M	Payoff for mating success by satellite and caller males
b_1, b_2	Background payoff opportunities for satellite males and females
F, $F\phi$	Costs to satellite and to caller in caller-satellite encounter without ambush and female absent
E	Cost to caller of giving an honest signal
A	Damage to an ambush victim

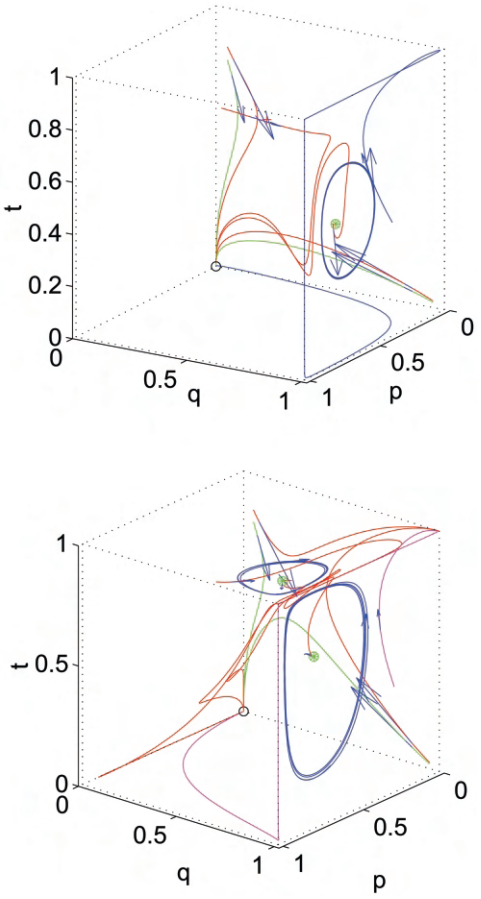

Figure 3: Strategy dynamics in the ambush game with mating readiness cost ($E > 0$). The collection of closed orbits on face $F1$ exists if $b_2 > N$ (*bottom*) but not if the opposite holds (*top*).

Results. There is a single determinant for the behavior of trajectories on the surface of the cube: whether a female prefers mating with a satellite male over her background opportunity (N vs. b_2). If $b_2 > N$, cycles are created on the face $t \equiv 1$. If $E > 0$ (fig. 3), the nontrivial equilibria always include a center on the face $q \equiv 1$ and a saddle on the back face $p \equiv 0$. There may also be an interior fixed point (or two, though this is extremely rare in parameter space). Also on the top face, a center is generated whenever the benefit from mating with a satellite male does not exceed the female listener's background opportunity. The complete breakdown equilibrium at the vertex (0, 0, 0) is always locally stable when $E > 0$. Usually, it is also the global attractor for all interior trajectories.

If the cost E for an honest but unsuccessful call is 0 (or no greater than the cost for a deceptive call), then the possibility for a long-term communicative pattern strengthens. Figure 4 shows strategy dynamics in which the edges (1, 1, ·) and (0, 0, ·), corresponding to complete belief or disbelief, are equilibria sets with both stable and unstable segments. They are the results of the center on $q \equiv 1$ and the saddle on $p \equiv 0$ transitioning off the signaling cube. The attractor that was situated at (0, 0, 0) has expanded to the bottom portion of the disbelief edge (0, 0, ·). The top portion of belief edge (1, 1, ·) is an attractor if the payoff to females for mating with a satellite is higher than her background opportunity. Otherwise, a center will appear on the top face $t \equiv 1$. The dynamics in the case of $E = 0$ are somewhat reminiscent of the two-man confidence game (app. D).

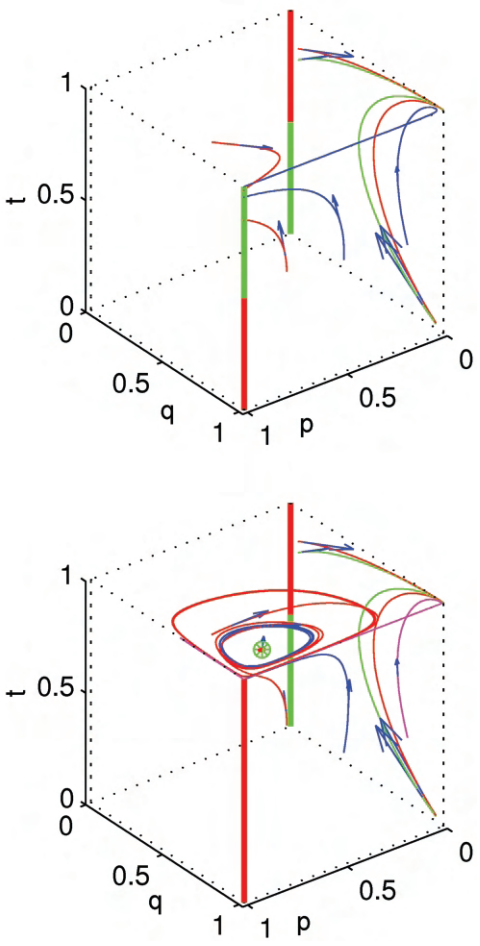

Figure 4: Examples of possible strategy dynamics in the ambush game without a mating readiness cost ($E = 0$). *Top*, female's payoff for mating with a satellite male is higher than her background opportunity; *bottom*, opposite holds. The front and back edges ($E9$ and $E11$) are lines of neutrally stable equilibria. Edge segments drawn in green are attracting, and those drawn in red are repelling. The circled asterisk in the top face of the cube indicates a fixed point.

For some combinations of parameters, using false mating signals to lure opponents into an ambush can be a successful strategy. The key factor is that satellite males and females exhibit differing levels of tolerance toward the costs of being injured. Female willingness to settle for matings with satellite males improves the likelihood that a partial ambush strategy will not cause females to avoid the territory. Therefore, a strategy that seems wrongheaded—ambushing blindly—can in fact serve to create a distribution of listener behaviors that benefits the signaler.

Partially deceptive mating calls cannot occur in either of the reduced two-player games (app. E). With females as the only listeners, the only possible stable equilibria are (0, 0) and (1, 1), and these attract all trajectories (except possibly the stable manifolds of unstable equilibria). At (1, 1) the mating call is always honest, and females always enter the territory in response. At (0, 0) the call is always dishonest (i.e., the caller intends to ambush), and females never enter. With satellite males as the only listeners, the only possible outcome is convergence to (0, 0). In either of the reduced games, at the (0, 0) equilibrium the signal has been become "inverted" into an honest signal of intent to ambush, which deters any listeners.

The logic of our ambush model may apply to more than the evolution of mating signals. For example, when an individual is signaling to attract any kind of future cooperator, there may be a risk that parasitic individuals (analogous to satellites) will try to steal the recruited cooperators for themselves. In such cases it might pay for signalers to employ an ambush to reduce the threat of parasitism, even if this may deter some potential cooperators.

Several important conclusions emerge from the ambush game. First, useful deception does not necessarily destabilize belief. In figure 4, the green segment of the front edge $E9$ is stable and able to resist some drifting of the level of honesty in the signals produced. Only if the neutral drift is of sufficient magnitude to push the equilibrium into the unstable region will selective pressure reverse for the listeners. A similar phenomenon can occur on the back edge.

Second, a signaler can benefit from using a mixture of honesty and dishonesty when faced with dispositionally distinct receivers. This is a principal mechanism for the maintenance of partial deception in our model. If the signal is "honest enough," then females should act as if it were true, so the signaler does not need to be totally honest. If the signal is "deceptive enough," then satellites should stay out, so total dishonesty is not necessary. Depending on parameters, there may be a range of deception rates that are both honest enough and deceptive enough, so a mixed strategy teases out differences between the receivers in the way that is most beneficial to the signaler. If deception is costly, signalers will prefer the lowest such deception rate. If deception is free, there may be a range of deception rates with equal net payoff to the signaler, leading to a line of neutrally stable equilibria in our model.

Finally, the ambush game demonstrates how a two-listener interaction can behave very differently from either of the reduced one-listener interactions. "Managing" the responses of two listeners with different payoffs may involve very different strategies than managing either listener alone.

Both predictions have important implications for field studies of communication. For example, territorial male songbirds frequently have two receivers for their songs: rival males attempting territorial intrusions and females attracted to the territory for mating (Bradbury and Vehrencamp 1998). Most field studies of the function of the song have focused on one kind of receiver. Evidence that songs of a given type have puzzlingly variable connection to subsequent singer behavior (as often reported) may indicate not nonadaptive "slop" in song functioning but rather a mixed strategy that yields fitness benefits to singers confronted with behaviorally distinct receivers. Empirical researchers should investigate this possibility in situations where signalers interact with multiple receivers in different roles.

Bluffing by Territory Holders

In this section we present a model for another signaling interaction involving territory-holding males: competition with other males for possession of the territory. Our previous example illustrated the potentially large difference between one-listener and two-listener interactions. This example illustrates the potential importance of considering strategy dynamics.

Calls by territory-holding males often play a role in competition with conspecifics over ownership of territories such as foraging areas or mating sites. The value of territories produces an inherent conflict between signalers and listeners. Therefore, we expect to see some manipulation of the signaling system. Male green frogs *Rana clamitans* produce acoustic signals as an indicator of size, but occasionally smaller males will lower their pitch so as to appear larger (Bee et al. 2000). Among crustaceans, many species of newly molted stomatopods such as *Gonodactylus bredini* (Steger and Caldwell 1983) engage in aggressive meral spread displays despite being physically unable to engage in combat. In both instances, challengers frequently would accept the signal as valid and would avoid direct conflict with the territory holder. The fiddler crab *Uca annulipes* also uses deceptive signaling in territory defense and in attracting mates (Backwell et al. 2000). If a fiddler

crab loses his enlarged brachychelous claw, the regenerated claw is leptochelous. That is, it is weaker because less muscle tissue is included in the regenerated limb, but it is also longer and lighter as a result. A crab with a leptochelous claw is at a disadvantage during combat but can engage in active display (waving) at a reduced energetic cost. The level of dishonest signaling in these populations can be quite high. Twenty percent of *G. bredini* are estimated to be molting deceivers, while among *U. annulipes* leptochelous deception can be employed by as much as 44% of the population.

As a minimal model for this type of interaction, we assume that signalers can be of two types: competitively superior (big) or competitively inferior (small). Likewise, listeners can be big (listener type 1) or small (listener type 2). Signalers produce a signal intended to mean "I am big, and this is my territory." If a listener believes that the signaler really is big, he will avoid a confrontation, but if the listener doubts that the signaler is big, he might challenge for control of the territory. A big male will always defeat a small male in combat, but conflicts between same-sized individuals will impose a higher penalty on both combatants because of the risk of escalating violence. There may also be an incumbency advantage for a territory-holding male in conflicts with an intruder of the same size.

Listeners have the choice to either respect the signal ($p, q = 1$) or challenge the resident for the territory ($p, q = 0$). In this model, either listeners are learning or the changing levels of belief reflect the evolving distribution of traits that determine respect or challenge behavior based on size. For signalers, big and small are not switchable roles—if they were, all individuals would choose to be big. Rather, the variable t represents the overall relative frequency of honest versus dishonest calls (i.e., calls from big vs. small territory holders) in the listeners' environment. For example, there may be fixed numbers of big and small individuals, but each of these can adjust its rate of signaling when it holds a territory, based on the payoff for doing so. As a result, regardless of the absolute numbers of big and small signalers, the fraction of signals that are honest (i.e., given by big signalers) may take any value between 0 and 1. Although our model is not derived from individual-level assumptions about the response of signaling rate to payoff, it has the right qualitative behavior: when the payoff to big signalers is higher than the payoff to small signalers, there is an increase in the big : small ratio among active signalers and vice versa. As usual the dynamics can be interpreted either as individual learning or as evolutionary change in traits determining the relationship between size and signaling rate.

Payoff Matrices. The payoff matrix for the territory challenge model is (with BL = big listener, SL = small listener)

Big (honest) Signaler

	SL: Respect	SL: Challenge
BL: Respect	b_1, b_2, T	$b_1, -F_3, T - F_3$
BL: Challenge	$\frac{T}{2} - F_1 - I, b_2, \frac{T}{2} + I - F_1$	$\frac{T}{2} - F_1 - I, 0, \frac{T}{2} + I - F_1$

Small (dishonest) Signaler

	SL: Respect	SL: Challenge
BL: Respect	$b_1, b_2, T + \epsilon$	$b_1, \frac{T+\epsilon}{2} - F_2 - I, \frac{T+\epsilon}{2} - F_2 + I$
BL: Challenge	$T - F_3, b_2, -F_3$	$T - F_3, 0, -F_3$

(9)

The rationale behind the payoffs is as follows (model parameters are summarized in table 2). First consider the case of a big signaler, meaning that the signal is honest. If the big and small listeners both respect the signal (first row, first column of the matrix), they receive their respective background reproductive payoffs, b_1 and b_2, and the signaler wins the territory of value T. If the big listener respects the signal but the small listener challenges (first row, second column), the big listener receives its background reproductive payoff, b_1, the small listener incurs only the fighting cost F_3, and the signaler wins the resource with net payoff $T - F_3$. If the big listener challenges but the small listener respects the signal (second row, first column), the big listener splits the value of the resource with the signaler but pays the fighting cost for size-matched males and the additional cost I for being a nonterritory holder, for a net benefit of $T/2 - F_1 - I$; the signaler therefore receives the net payoff $T/2 - F_1 + I$, and the small listener receives its background reproductive payoff, b_2. If the big and small listeners both challenge (second row, second column), the big listener again receives a net benefit of $T/2 - F_1 - I$, and the signaler again receives the net payoff $T/2 - F_1 + I$, but the small listener is overwhelmed and receives zero payoff. We assume that $F_3 < \min(F_1, F_2)$ because of the tendency for aggression to escalate when opponents are similar in size.

Next consider the case of a small signaler, so the signal is dishonest (bottom matrix). If the big and small listeners both respect the signal (first row, first column), they receive their respective background reproductive payoffs, b_1 and b_2, and the signaler wins the territory of value $T + \epsilon$. A value of $\epsilon > 0$ means that small males have a higher payoff from territory holding, perhaps through a trade-off between gonad development and fighting capability (Tomkins and Simmons 2002) or a reduced cost of signal production as in leptochelous fiddler crabs; we refer to

Table 2: Parameters of the territory challenge game

Parameter	Description
b_1, b_2	Background payoff opportunities for big and small listeners
T	Value of the territory
ϵ	Possible greater benefit from territory to small males
F_1, F_2	Cost of fighting a same-size individual for big and small males
I	Advantage of the territory holder in fights between same-size males
F_3	Cost of a fight between different-size males, $F_3 < \min(F_1, F_2)$

this as a territory bonus to small males. If the big listener respects the signal but the small listener challenges (first row, second column), the big listener receives its background reproductive payoff, b_1, the small listener splits the territory value but incurs the fighting cost F_2 and suffers the additional cost I for being a nonterritory owner, for net payoff $(T + \epsilon)/2 - F_2 - I$, and the signaler consequently receives the payoff $(T + \epsilon)/2 - F_2 + I$. If the big listener challenges but the small listener respects the signal (second row, first column), the big listener wins the resource with net value $T - F_3$, the small listener receives its background reproductive payoff, b_2, and the signaler incurs only the fighting cost F_3. Finally, if the big and small listeners both challenge (second row, second column), the big listener again receives a net benefit of $T - F_3$, and the signaler again incurs only the fighting cost F_3, but the small listener is overwhelmed and receives zero payoff.

If there is no territory bonus for a small territory holder ($\epsilon = 0$), then selection on signaler honesty is neutral when both listener types respect the signal (edge $E9$ or $p = q = 1$). When there is a territory bonus, the rate along this edge is negative because small males make better use of a territory.

Several limits on the parameters can be assumed for this model. (1) When facing a big challenger, a territory holder is better off being big: $T/2 + F_3 + I > F_1$. (2) When facing a big defender, a big listener is better off not challenging: $T/2 < b_1 + F_1 + I$. (3) When facing a small defender, a big listener benefits more from challenging: $T > B_1 + F_3$.

Big listeners therefore should challenge a signaler when most territory holders are small, but they want to avoid challenging big territory holders. We do not exclude the possibility that a small listener would gain from challenging a big territory holder because of the possible territory bonus and the lower cost of fights between different-size individuals. We also allow parameters such that small listeners never gain from challenging a territory holder of any size.

Results. Our assumptions limit the possible equilibria in this game. Besides the eight vertices, there is always a center on the face $q \equiv 1$. The set of periodic orbits around this center typically serves as the attractor. In addition there is also the possibility of centers appearing on the faces $q \equiv 0$ or $p \equiv 1$. These points and the sets of surrounding closed orbits can be attractors or repellors. When there is a double attractor, the boundary between the basins of attraction includes a spiral fixed point. When the territory bonus ϵ is absent, the center on the face $q \equiv 1$ shifts to the edge $(p, q, t) = (1, 1, \cdot)$, thus creating a new edge equilibrium, with the upper half of the edge being an attractor. Other centers on faces of the cube are not affected.

A particularly interesting example of this model's dynamics is shown in figure 5. This result was produced by a relatively large bonus for small territory holders. There are two distinct sets of attracting cycles, one where honesty oscillates with the belief level of big listeners and the other where honesty and the belief of small listeners cycle. In each case the other type of listener assumes that all signals are truthful. Thus, for the trajectories that go toward the cycles with $p \equiv 1$, small listeners are actually more daring than big listeners. Figure 5*d* shows a numerical calculation of the boundary between the two attractors.

In figure 6 the complete range of possibilities is shown. There are three main factors determining the dynamics on the surface of the cube. The first is whether there is a territory bonus for small males. The second is the comparison between $(T - \epsilon)/2 + (F_2 - F_3)$ and I. This is the relative cost of holding on to a territory when facing challenges from small males. If the former is greater, then flow along the edge $(p, q, t) = (1, 0, \cdot)$ is positive (first and third rows of plots in fig. 6). If the latter is greater, then flow is downward (second and fourth rows). The final determinant is whether it is profitable for a small listener to challenge a small signaler. If not, then small individuals will always assume that a signal is truthful (fig. 6, *left column*). Otherwise small challengers will attack small defenders (fig. 6, *right column*).

When the territory bonus to small males is absent, the top half of the edge $(1, 1, \cdot)$ is always globally attracting. Although there is partially neutral drift of honesty, if the amount of lying becomes too large, the system will self-correct and return to a higher rate of honest signaling. The bottom two rows of figure 6 demonstrate how the dynamics of the system are altered when ϵ is eliminated.

What we see in this model is that some nontrivial level

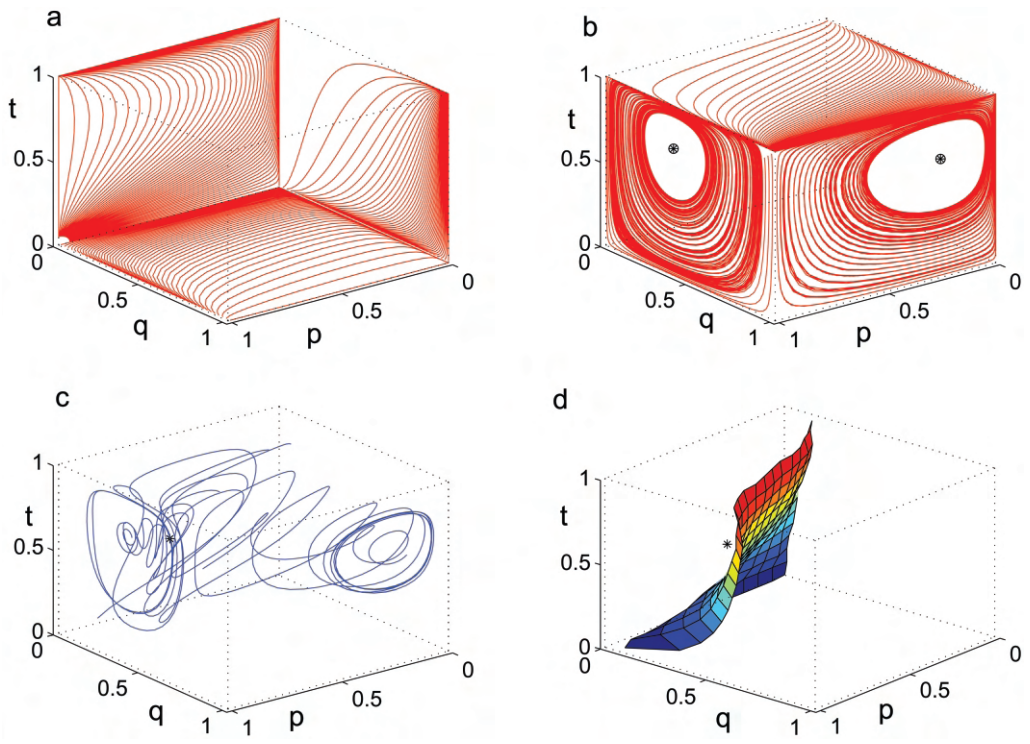

Figure 5: Coexisting attractors in the territory challenge game when there is a large territory bonus for small listeners. There are two basins of attraction: one converging to big listeners always respecting the signal and one converging to small listeners always respecting it. *a*, *b*, Trajectories on the faces of the signal cube. The two sets of periodic orbits in *b* are both attractors. *c*, Two trajectories that start in the interior and are attracted onto periodic orbits on the cube surface. The colored surface in *d* is the boundary between the basins of attraction for the sets of periodic orbits. Fixed points are shown as asterisks. Parameter values are $T = 10$, $\epsilon = 6$, $b_1 = 3$, $b_2 = 1$, $F_2 = 3$, $F_1 = 4$, and $I = 1$.

of deception is always present. Small males can effectively establish themselves as territory holders because the risk of fighting a big defender is a deterrent against challengers. This deception can occur either at a static level, such as when $\epsilon = 0$, or in the context of a cycling pattern with one of the two listeners' belief level.

In contrast to our other examples, the territory challenge game has the property that the one-listener games are both nested within the two-listener game. Specifically, on the faces $p \equiv 1$ and $q \equiv 1$, big and small listeners, respectively, are constrained not to challenge a territory holder—which is equivalent to removing that type of listener from this game. These faces are not necessarily attractors for the two-listener game, so the one-listener games still require a separate analysis. However, the conclusions from this analysis (see app. F) are not surprising: depending on parameter values, there may be complete breakdown, complete belief, or stable partial deception on either point equilibria or periodic trajectories.

The contrast between two-listener and one-listener games is therefore less severe in this example than in our others, in that one-listener reduced games can exhibit stable partial deception. However, the dynamic game structure is an essential feature. Stable partial deception cannot be a classical ESS solution in any one-listener (nonrepeated) matrix game (this follows from Selten's theorem for bimatrix games), but it occurs here in both one-listener dynamic models. In the two-listener game, the results in appendix B show that the cases with periodic orbit attractors also do not have any static ESS solutions. One-listener models do miss the transient behavior of the system when all three players begin with mixed strategies. In particular, the double oscillator attractor represented in figure 5 cannot occur without the inclusion of all three players. Discovering the full range of possible outcomes thus also depends on recognizing that listeners are heterogeneous, which would be inevitable in territorial advertising so long as potential challengers differ in fighting ability.

Thus, the territory challenge game reveals that multiple listeners can increase the complexity of signaling strategy dynamics in ways that can have implications for field and laboratory studies. The predicted enrichment of the dynamics (e.g., the double oscillator attractor) might be

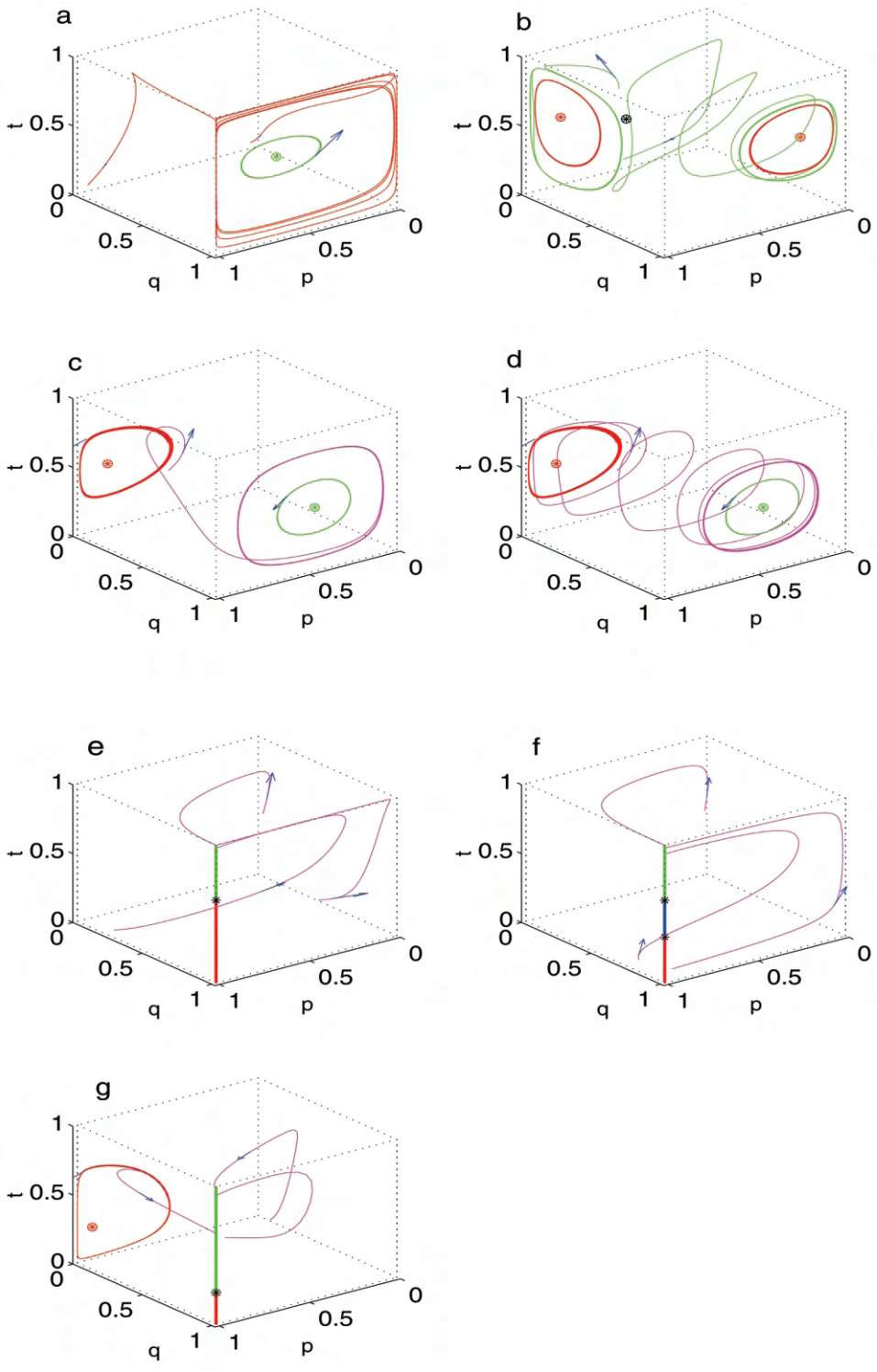

Figure 6: Possible dynamics in the territory challenge game. *a–d*, Possible behaviors when there is a territory bonus for small individuals. *e–g*, Corresponding dynamics when this bonus is removed; parameter conditions corresponding to *d* cannot occur without a territory bonus for small individuals.

tested directly in cases in which learning is involved—for example, signal honesty could be tracked over time in laboratory studies of territorial interactions among interactants deployed as pairs versus triplets. A final point is that the multiple-listener model may have important theoretical implications for the likelihood of invasion by assessor strategies (i.e., direct but costly assessment of opponents' resource holding power instead of sole reliance on signals). Specifically, if the existence of multiple listeners reduces the mean level of signal honesty or leads to oscillations during which the frequency of honesty is sometimes quite low, assessment strategies may be more likely to invade.

Discussion

Animals communicate dishonestly with one another over food, territory, and mates. If a given communication system can be gamed, individuals will arise who do so, yet populations continue to accept various signals as truthful. Signaling systems that cannot endure the stress of false communication will fail or else they have already failed. Observed instances of animal communication have presumably withstood a history of attempts to introduce deceptive elements and either excluded or accommodated dishonesty.

The issue of truthfulness in animal signals has had a storied history. Originally viewed as an inherently honest interaction between cooperative individuals, signaling came under great scrutiny with the introduction of game theory to behavioral ecology. Because of the potential benefits to deceitful mutants, honest signaling was believed to be evolutionarily unsustainable except in circumstances where deception was impossible (e.g., body size as a signal of fighting ability). Zahavi's handicap principle then provided a mechanism for stable maintenance of honest signals. Ideas such as transiency, external noise, and verification costs have been advanced to explain the persistent presence of dishonest signals.

In this article we have shown that signaling in bad faith can be a persistent outcome of a signaling interaction as a natural result of the payoffs to the parties involved. Within our general framework, simple models inspired by real conflict-of-interest signaling problems have repeatedly shown a tendency to produce appreciable levels of dishonesty without jeopardizing the overall occurrence of communication.

Our model uses payoff matrices as the basis for a differential equation model that describes the evolution of the levels of honesty and belief. This approach has some similarities to the classical replicator equations model (Hofbauer and Sigmund 1998), which describes competition among a finite set of haploid clones, each clone having a particular static (pure or mixed) strategy. However, our model is more similar to quantitative genetics models for selection response. Alternatively, the model can be interpreted as learning and behavioral change within an individual's lifetime, with the $x(1 - x)$ terms representing encounter rates between individuals playing different strategies, hence the rate at which individuals gain information about the potential benefit of a change in strategy.

The dynamic model could exhibit a variety of behaviors and equilibria. The vertices represent populations playing pure strategies. Other possible types of equilibrium sets include the external edges and faces, saddles, limit cycles, centers on the faces, internal spirals, and lines and curves. There are restrictions on the combinations of features that can be produced at any one time (Rowell 2003), but there is a fascinating variety of possible long-term and transient behaviors. Each of these results reinforces the possibility for nontrivial but nondestabilizing amounts of deception.

Directions for Future Theoretical Research

An important direction for future research is to see how sensitive predictions are to the form of the dynamic model. In many cases the attractors in our specific models are structurally unstable—not robust against generic perturbations of the underlying dynamical system. This structural instability derives from our assumption of random pairwise interactions and constant payoffs, leading to a matrix game structure for payoffs. Many influential game models (such as hawk-dove) have been matrix games because their simplicity made them a good vehicle for introducing new ideas that could then be generalized and extended. Our model here is in the same spirit. A mating call by a territory-holding male gets him a mating, a fight, or no response at all, depending on the current strategy mix of listeners (as influenced by the past strategy mix of signalers). For a first look at the impact of third-party listeners, it seems reasonable to posit listeners drawn at random from the population, constant payoffs per mating, and a constant size-dependent cost for a fight. In reality a recently bruised satellite might be shy of the next call, the fitness gain from an additional mating will depend on the overall mating rate, and a satellite who has been ambushed before might be better at fighting back. Qualitative properties ought to be robust against such complications—such as the potential benefits from a signal that is honest enough for some listeners to believe and deceptive enough for others to disbelieve—but this remains to be seen.

Another important next step is to expand the signaler's options. For simplicity we used the simplest possible structure: a single signal whose veracity must be judged by the

listener. For many applications this is too simple. At a minimum, a signaler has the option of remaining silent rather than giving a signal. More generally (though mathematically equivalent) there may be several types of signal—different alarm calls for different types of predator (reviewed by Searcy and Nowicki 2005), different intensities of begging to indicate need, different levels of a quality advertisement—and the signals may differ in reliability.

For simplicity our model is formulated at the population level and ignores any individual differences other than their role in the game. We also ignore finite population effects—the equivalent of demographic stochasticity—in particular differences among individuals in their past experience and consequently their beliefs about population composition and relative payoffs. One can imagine a range of models of increasing complexity, starting from ours and terminating in individual-based simulations of agents whose learning rules, as well as their behaviors, could change over time as individuals interact and reproduce at rates that depend on their payoffs.

Finally, it would be interesting to consider extensions of our model to situations where players can recognize each other and build up a history of interactions with each other. Conditional strategies such as "once bitten twice shy" (disbelieve somebody who lied to you before, analogous to tit for tat) might then be advantageous. This extension would require tractable models for strategy dynamics in the space of conditional strategies, which could then be contrasted to classical solution concepts for repeated games.

Possible Experimental Tests

Finally, we discuss some potential experimental scenarios for testing our model and the conclusions that we have derived. The two-man confidence game described in appendix D is well suited for implementation in an experiment with human subjects. In this game a signaler is trying to aid one listener (listener 2) and deceive another (listener 1). All three participants are aware of the nature of this game and the payoff matrix. Divide the study participants into three pools. The participants will be isolated from one another at terminals and told their role (signaler, listener 1, or listener 2). At the terminal, they will record whether they play truth or lie (if a signaler) or belief or disbelief (if a listener). Thus, for instance, begin with 10 participants in each of the three pools, and allow a computer to randomly assign which set of three interacts during each round. Therefore, 10 games would be occurring simultaneously. At the end of each round, each player would be informed of the strategy frequencies employed in that round. Knowing the game, players can then infer whether they would have been better off playing another strategy. Consequently learning and behavioral adaptation can occur without any explicit mechanism for exchange of information between players. This experiment can test three qualitative predictions of the confidence game model. (1) Outsider listeners (listener 1) will mimic the current strategy preferences of favored listeners (listener 2). (2) Listeners will tend toward a long-term pure strategy. (3) Signalers will continue to randomize the level of honesty in their communication.

The same kind of setup with human subjects could actually be used to implement any set of nonnegative payoff matrices, using monetary rewards. Because only relative payoffs matter in the model, the payoffs for any game can be modified by adding a constant to each, to make all entries nonnegative. Consequently, experiments with human subjects can be used in principle to test the predictions from any specific model, even if the payoff matrices are motivated by a signaling interaction between nonhuman animals. Monitoring how individuals' behavior changes over time will be a strong test of the model because both the transient strategy dynamics and the long-term outcomes can be compared with model predictions.

Games involving public signals might also be testable with animal subjects, using food items for the payoffs and modified payoff matrices with nonnegative entries. Experimenters play the role of the signaler, producing public signals to which listening individuals can respond by entering an area with a feeder. Listeners would be as similar as possible in their inherent properties, but each would arbitrarily be assigned to each of the roles (e.g., satellite male and female). The payoff dispensed by the feeder would then correspond to the appropriate payoff, given the nature of the signal and the response of the listeners (e.g., no food for a satellite male who entered alone in response to a deceptive mating call). We suggest two different possibilities for the dynamics of signaler truthfulness. First, the experimenters could dynamically adjust their frequency of true and false signals as dictated by the model; in this situation, only the listener component of the model would be tested. Second, the signaler could be a human subject aware of the payoff matrix and receiving monetary payoffs based on the responses of the animal subjects, as per the payoff matrix of the game.

In our ambush game, the presence of intercepting satellite males can dramatically change the outcome from strictly honest mate calling to occasional deceptive (ambush) calling, resulting in male aggression against potential mates. Few studies have examined the possibility of such deception, probably because of the assumption that males would never have an incentive to provide a false mating signal that could potentially lead to loss of, or injury to, a potential mate. However, predictions of the ambush hy-

pothesis should be amenable to experimental test. For example, under the caller-ambush hypothesis, male aggression toward females should be most intense when satellite males occur at relatively high frequencies, as in fish with territorial male/satellite male dimorphisms.

Acknowledgments

This article is based on work described in the doctoral dissertation "Diffusive Food Webs and Signaling Dynamics of Populations" (Rowell 2003), submitted in partial fulfillment of the requirements for the PhD in applied mathematics at Cornell University. We thank the other advisory committee members (J. Guckenheimer and N. G. Hairston Jr.) for their advice and comments on the dissertation and P. Hurtado, L. Jones, V. Pasour, and three anonymous reviewers for helpful comments on drafts of the manuscript.

APPENDIX A

Equilibria in the Signaling Cube

In this appendix we briefly discuss some of the equilibria that can occur in the three-party signaling model. Rowell (2003) gives a more comprehensive explanation of these and other results, as well as proofs for the results in this appendix.

System dynamics occur within the signaling cube (fig. 2). For discussing the stability of various equilibrium structures, we will refer to elements of the Jacobian matrix

$$\frac{\partial(p', q', t')}{\partial(p, q, t)} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}. \tag{A1}$$

The eigenvalues of the Jacobian matrix satisfy the third-order characteristic equation

$$-\lambda^3 + \lambda^2(a + e + i) + \lambda(fh + bd + cg - ei - ai - ae) + (aei + bfg + cdh - afh - bdi - ceg) = 0. \tag{A2}$$

The different equilibria that might occur as a result of the model dynamics include (1) the eight vertices of the cube, (2) individual cube edges, (3) isolated points or lines on an external face, (4) an entire face of the cube, and (5) isolated points or lines in the interior of the signaling cube.

The eight vertices of the signaling cube are always equilibria. The eigenvalues are always the trace elements a, e, and i. The corresponding eigenvectors are the orthonormal vectors **i**, **j**, and **k**, respectively. A vertex is a local sink if and only if it is strict Nash and hence an evolutionarily stable stragety (ESS) in the static interpretation. An asymptotically stable vertex is therefore a biologically meaningful steady state in which the extremal strategies (pure strategies or pure populations) are evolutionarily stable. A vertex is weak Nash if there is no outward flow along any connecting edge, but it need not be stable relative to completely mixed strategies.

An entire edge can serve as a set of equilibria when there is no one best response for a player faced with opponents using pure strategies. The player's two pure strategies are equally successful in the existing environment, so there is no directional change driven by selection. When there is an edge that is an equilibrium set, the interpretation is that only neutral drift produced by stochastic effects causes any change in that player's strategy. The eigenvalues for an edge equilibrium are $\{a, e, 0\}$, $\{a, 0, i\}$, or $\{0, e, i\}$, depending on whether the free variable is t, q, or p, respectively.

The following theorem sums up edge equilibria.

Theorem 1. If two players use pure strategies, then either the best response for the third player is a pure strategy or else all mixed strategies are equally successful.

External faces of the signaling cube may contain isolated fixed points. Saddles arise when there are two vertices that serve as local attractors (in opposition). Additionally, there may be a center surrounded by a series of closed orbits within the face. When t is extremal (0 or 1), any isolated fixed point on the faces $F1$ and $F2$ have eigenvalues i and $\pm(bd)^{1/2}$. If p or q is extremal, the eigenvalues are $\{a, \pm (fh)^{1/2}\}$ and $\{e, \pm (cg)^{1/2}\}$.

A line of equilibria may be created within a face if two opposing edges of a face are themselves equilibrium structures. Mathematically, this is a consequence of a rate function possessing a special linear polynomial factorization. For instance, if $(t - 1)$ is a factor of $r_p(t, q)$, then a line perpendicular to $(\cdot, 1, 1)$ and $(\cdot, 0, 1)$ might appear as an equilibrium structure on the face $t \equiv 1$. If that linear factor is common to two rate functions, then an entire face is in equilibrium.

The eigenvalues for equilibrium lines and planes on the external faces of the signaling cube are $\{a, 0, 0\}$, $\{e, 0, 0\}$, or $\{i, 0, 0\}$ for p, q, or t extremal, respectively. The double eigenvalue 0 is degenerate for lines but not for plane equilibria.

The following four theorems concern equilibria on faces in three-player games and equilibria in two-player games.

Theorem 2. In a two-player game, if there is selective pressure on one player when the other player uses pure strategy **X** but none when the other player uses $1 - \mathbf{X}$,

then the dynamical model admits no mixed strategy equilibrium.

Theorem 3. If two opposing edges of a face, with variable $\mathbf{X} \equiv x$, are equilibrium sets and at least one of the other edges has nontrivial flow, then the rate function for the state variable that varies over the static edges is reducible with factor $(\mathbf{X} - x)$. Furthermore, there exists neither an isolated equilibrium line parallel to these edges nor an isolated fixed point.

Theorem 4. In a two-player game, if one player is under no selective pressure, regardless of the other player's strategy, then there is an isolated equilibria solution if and only if the first player can induce different directions of selection pressure on the second player with different pure strategies.

Theorem 5. If an entire face is an equilibrium set and the rate functions are not trivial, then the opposing face of the signaling cube may have at most a single isolated fixed point.

Finally, there may also be attractors in the interior of the signaling cube. Multiple isolated fixed points and invariant curves may arise as a result of the dynamics of the model. Lines and curves are the result of special critical value bifurcation events. The key result for interior fixed points, that no such point is both stable and hyperbolic, is included as part of theorem 8 in this article.

APPENDIX B

Nash Equilibria, Evolutionarily Stable Strategies, and Dynamic Equilibria

A Nash equilibrium exists when an individual cannot do better by changing his strategy independently, given the actions of the other participants; thus, any evolutionarily stable strategy (ESS) is necessarily a Nash equilibrium. This may be a local Nash condition (only minor variations of strategy considered) or a global Nash condition (all variations of strategy considered), depending on the context and the particular game. For this article, it is the local sense of Nash that is important.

In our model the payoff for mixed strategies is computed by linear interpolation between the pure strategy payoffs. However, the relationship between Nash and dynamic equilibria still holds under the more general model (eq. [2]) with nonlinear payoffs. The following three theorems may be stated for this model; proofs are given at the end of this appendix. The importance of these results is that static equilibria and their properties can be deduced immediately from the fixed points of the dynamic model and their local stability properties.

Theorem 6. All Nash equilibria **s** to the extended game defined by the payoff structure $[f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})]$ are also fixed points of the dynamic model.

Theorem 7. The strategy $\mathbf{s} = (s_1, s_2, s_3)$ is a weak Nash equilibrium among pure strategies to the normal form linear game if and only if the vertex **s** has no unstable manifold, though it need not be locally stable. The strategy is a strict Nash, and hence a local ESS, if and only if the vertex is a hyperbolic, locally asymptotically stable sink.

A vertex that corresponds to a weak Nash strategy may be unstable relative to perturbations to the interior of the signaling cube. Consider the system

$$\begin{aligned} t' &= [-pq + q - 1]t(1 - t), \\ p' &= [t + q - 2]p(1 - p), \\ q' &= [-pt + t - 1]q(1 - q), \end{aligned} \tag{B1}$$

(this example is derived from a payoff matrix but does not have a biological interpretation). The pure-strategy vertex at (0, 1, 1) is weak Nash in that no individual player can improve his or her payoff by unilaterally switching strategy. However, any perturbation to the interior will initiate a move away from the vertex.

Theorem 8. Any Nash equilibrium that contains a mixed strategy for all three players in a signaling game whose mixed-strategy payoffs are linearly interpolated is weak. The corresponding fixed point of the system of differential equations is either unstable or nonhyperbolic.

Moreover, an isolated fixed point on any face of the signaling cube that is not at a vertex can be either weak Nash or not Nash at all. The non-Nash situation occurs whenever the fixed point has an unstable eigenvector pointing into the cube. For example, at a fixed point with $p = 0$, the solution flowing into the cube along the unstable manifold has $p' > 0$ and, hence, $\partial f_p/\partial p > 0$ near the fixed point. In general, for the player whose decision variable is extremal (0 or 1) on the face, the solution on the unstable manifold implies that the player's fitness can be improved by moving away from the equilibrium.

Proof of theorem 6. Let **s** be a Nash equilibrium; then $\partial f_i/\partial x_i|_{\mathbf{s}} = 0$ for each x_i that is nonextremal. But under the dynamic adaptive model used here, this implies that the variable x_i is in equilibrium. Because this is true for all nonextremal state variables and because all extremal variables are already static because of the logistic model, the strategy **s** is a fixed point in the dynamic system.

Proof of theorem 7. To be considered a weak Nash equilibrium, no individual i may improve his or her payoff by unilaterally switching his or her strategy to $\sim s_i$. That is, $f_i(s_i, s_{-i}) - f_i(\sim s_i, s_{-i}) \geq 0$. The left-hand side of the inequality is the differential payoff (up to sign because of the extremal switching of sign) used as the logistic rate for the ith equation in the dynamic model. Therefore, there

is no unstable flow away from the vertex along any of the three connecting edges. However, for any vertex of the dynamic game, the eigenvectors are the standard orthonormal unit vectors, and the eigenvalues are the logistic rates of growth (again up to sign). Therefore, the vertex has no positive eigenvalue.

If the inequality stated above is strict for each of the three players, then that vertex possesses no zero eigenvalue, either. Consequently, the fixed point is both hyperbolic and a stable sink.

Proof of theorem 8. Because the payoff for player i using a mixed strategy is determined as a linear interpolation between the payoffs accrued from playing the two pure strategies available, a mixed strategy is Nash only if the payoff differential is 0. Consequently, no player using such a strategy can improve upon his or her payoff by changing his or her strategy, but neither could such a player reduce his or her payoff by doing likewise. The strategy is therefore only weakly Nash. According to the dynamic model, all such points would have 0 for the diagonal elements of the Jacobian matrix.

Let **x** be an interior isolated fixed point. The eigenvalue equation for such a point is

$$-\lambda^3 + \lambda(fh + bd + cg) + (bfg + cdh) = 0. \quad \text{(B2)}$$

Some of the above remaining terms may be 0, depending on the exact type of point at the equilibrium, but the sum of all eigenvalues is equal to the trace of the Jacobian, or the coefficient for λ^2, which is 0. The roots of a third-order polynomial with real coefficients is either three real roots or one real root and one complex conjugate pair. If you assume that the point is hyperbolic, there is a real nonzero solution z to the polynomial equation. If $z > 0$, then the point has an unstable manifold, and the point is not stable. If $z < 0$, then at least one of the other two roots must have a positive real component. But if any real component of an eigenvalue is positive, then the point is again unstable. Therefore, if the point is hyperbolic, it cannot also be stable. The converse holds as well, logically.

APPENDIX C

One-Listener Games

In this appendix we consider the reduced two-player game with a single listener or type of listener and a signaler that potentially stands to gain from deceiving the listener. The parameters of the model are the payoffs to each player in the four possible situations: signaler truthful or not and listener believing or not:

	Truthful ($t = 1$)	Not Truthful ($t = 0$)
Believe ($p = 1$)	a_1, c_1	a_2, c_2
Not Believe ($p = 0$)	a_3, c_3	a_4, c_4

(C1)

where the a_i are the payoffs to the listener and the c_i are the payoffs to the signaler.

The dynamic model on $[0, 1] \times [0, 1]$ corresponding to the payoff matrix (eq. [C1]) is

$$t' = t(1 - t)[(c_1 - c_2)p + (c_3 - c_4)(1 - p)],$$
$$p' = p(1 - p)[(a_1 - a_3)t + (a_2 - a_4)(1 - t)], \quad \text{(C2)}$$

which we write for convenience as

$$t' = t(1 - t)[\alpha p + \beta(1 - p)],$$
$$p' = p(1 - p)[\gamma t + \delta(1 - t)]. \quad \text{(C3)}$$

The corners of the signaling square are always fixed points. Apart from the nongeneric situation where one of the coefficients in equation (C3) is exactly 0, there is an interior fixed point if and only if α and β have opposite signs and also γ and δ have opposite signs. When an interior equilibrium exists, the Jacobian there has zeros on the diagonal; hence its trace is 0. The determinant of the Jacobian has the sign of $-(\alpha - \beta)(\gamma - \delta)$, which is the same as the sign of $-\alpha\gamma$ due to the condition for the fixed point to exist. An interior equilibrium is therefore a saddle if $\alpha\gamma > 0$ and a center if $\alpha\gamma < 0$, whose stability is not determined by linear stability analysis.

The assumed nature of the signaling interaction can be expressed as conditions about the payoffs. If the interests of the signaler and listener are opposed, then for the listener it is best to believe a truthful signal and disbelieve a deceptive signal:

$$a_1 > a_3,$$
$$a_2 < a_4. \quad \text{(C4)}$$

Similarly, when the listener believes the signaler gains by lying,

$$c_1 < c_2. \quad \text{(C5)}$$

These conditions are really the definition of truthful and deceptive signals.

If the listener always disbelieves, then the signaler may still gain by lying ($c_3 < c_4$). For example, the signaler may benefit from any occasions when listeners accidentally depart from disbelief. Or, the signaler may do better by telling the truth and avoiding a cost of deception ($c_3 > c_4$).

If $c_3 \leq c_4$, then equation (C5) implies that $t' < 0$ every-

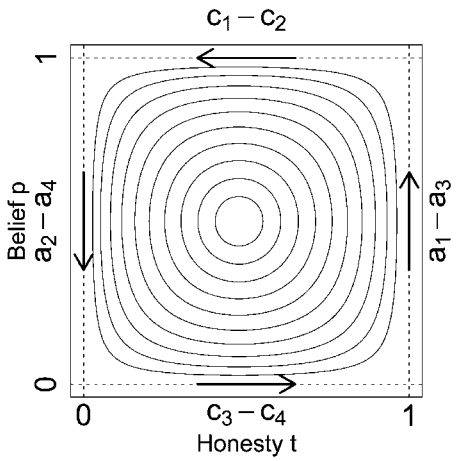

Figure C1: Strategy dynamics in the general two-player game. The expression next to each edge is the parameter combination whose sign gives the direction of flow along that edge. The arrows indicate the direction of flow under the assumptions stated in the text for a one-listener game with $c_3 > c_4$. Numerical solutions of the dynamic model are plotted for a set of interior initial conditions.

where in the interior of the square. As t decreases, equation (C4) implies that eventually $p' < 0$. Nullcline analysis confirms that the system converges to complete breakdown: the signaler always lies ($t = 0$), and the listener always disbelieves ($p = 0$). The only possibility for deception without complete breakdown is therefore when

$$c_3 > c_4. \tag{C6}$$

In this case the dynamics are inherently cyclic (fig. C1). The directions of flow on the boundary of the square are counterclockwise rotation. The eigenvalues at the interior fixed point are always pure imaginary, suggesting a center. As in the Volterra-Lotka predator-prey equations, separation of variables can be used to find a first integral for the system. This proves that the interior is a center, surrounded at least locally by a family of neutrally stable closed periodic orbits. Numerical solutions suggest that the square is completely filled by periodic orbits. The model thus supports persistent deception without breakdown of belief when $c_3 > c_4$, but it does not support stable deception.

The avoidance of breakdown with a single listener type therefore requires $c_3 > c_4$: when the degree of belief is low, a dishonest signal is less beneficial for the signaler than an honest one. One way that this could arise (though not the only way) is if there is an intrinsic cost to giving a dishonest signal. Intrinsic cost means that the signal cost is "charged" when the signal is issued rather than a result of listener responses. This is similar to the handicap mechanism in that an intrinsic cost of deception keeps the signaler from being so dishonest that listeners all believe.

In a similar way, a second listener type can prevent complete breakdown by behaving in a way that penalizes the signaler when the frequency of deception is too high. The cost of deception is then a frequency-dependent emergent result of the behavioral dynamics resulting from the signaling interaction.

We can compare the dynamic model to a static game with the same payoffs. If $c_3 < c_4$, then the signaler always gets a higher payoff from dishonesty and should always lie. The listener then should always disbelieve, so the one and only evolutionarily stable strategy (ESS) is complete breakdown: $t = 0$, $p = 0$. For $c_3 > c_4$ the interior fixed point of the dynamic model is a Nash equilibrium of the static game; that is, a unilateral deviation by either player is payoff neutral. However, by Selten's theorem any ESS of a bimatrix game must consist of pure strategies (Hofbauer and Sigmund 1988, p. 138), so the interior fixed point is not an ESS.

APPENDIX D

Two-Man Confidence Game

As a second paradigm for the role of dishonesty in a third-party listener situation, we consider an idealized two-man confidence game. Two listeners engage in a zero-sum symmetric wagering game on the validity of a signal produced by the signaler. The signaler, however, actually has a vested interest in assuring that the second listener wins the contest. This may be due to a kickback of a small amount each time the second listener wins, or it may be a psy-

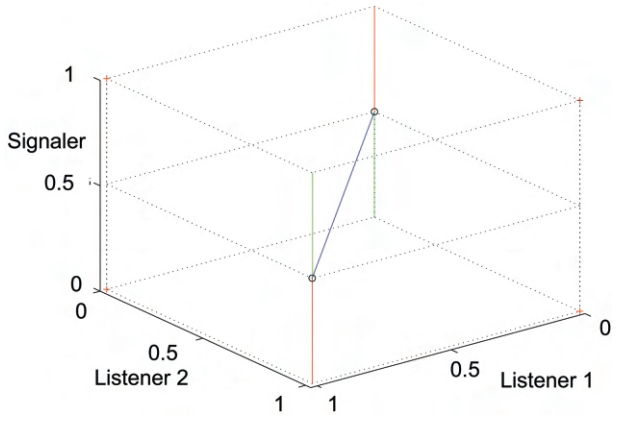

Figure D1: Equilibria for the two-man confidence game. The equilibria consist of the vertices, the front edge, the back edge, and an interior line. Circles represent the intersection of the interior line with the edges as well as the transition between attracting and repelling regions.

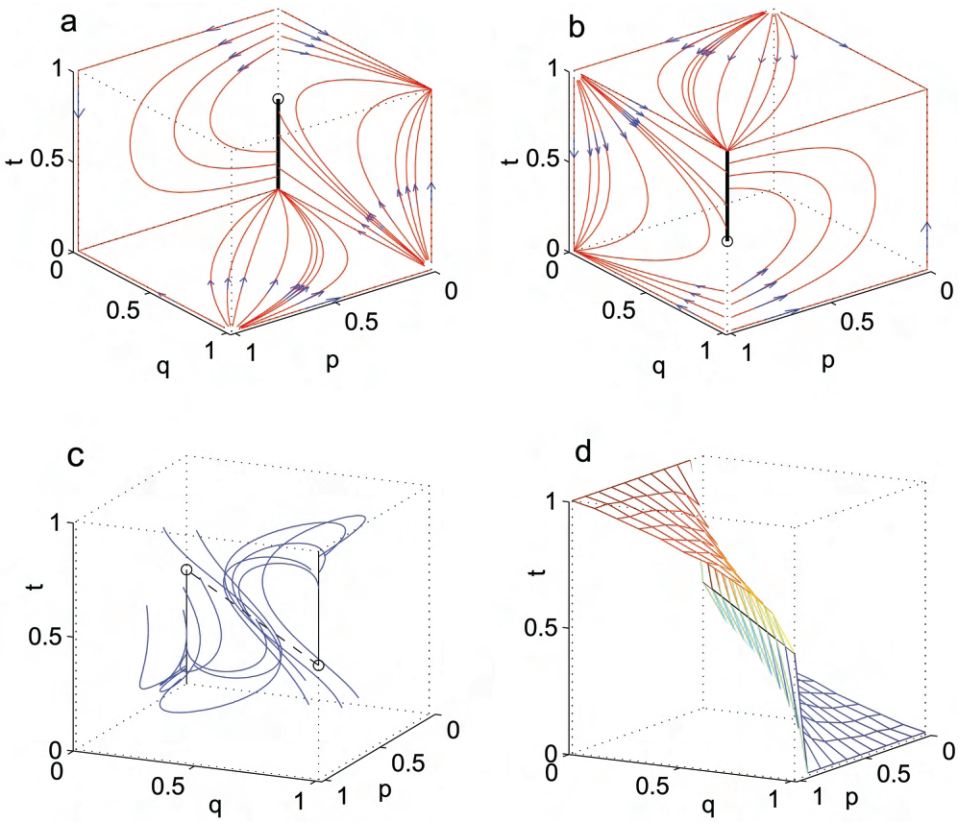

Figure D2: Strategy dynamics in the two-man confidence game. *a*, Surface trajectories in the back and bottom faces. *b*, Surface trajectories in the front and top faces. *c*, Interior trajectories and the two attracting regions. *d*, Boundary between the basins of attraction.

chological benefit or a benefit of some other kind. The complication is that there is no way for the signaler to give his accomplice a surreptitious secondary signal to indicate the validity of the public signal. Furthermore, the two partners are prohibited from prearranging a pattern of honest and dishonest signals. This game is somewhat atypical because of the symmetry between true and false signals, but it illustrates some important aspects of our model in a simple context and leads to some predictions that may be useful for experimental tests as described in "Discussion."

The payoff matrices corresponding to our assumptions are

	$t=1$		$t=0$	
$p=1$	$0,0,0$	$W,-W,0$	$0,0,0$	$-W,W,\epsilon$
$p=0$	$-W,W,\epsilon$	$0,0,0$	$W,-W,0$	$0,0,0$
	$q=1$	$q=0$	$q=1$	$q=0$

(D1)

where W is the wager cost and ϵ is the payoff to the signaler when his partner (listener 2) wins. The resulting dynamic model is

$$\begin{aligned} t' &= \epsilon(q-p)t(1-t), \\ p' &= W(2t-1)p(1-p), \\ q' &= W(2t-1)q(1-q). \end{aligned} \tag{D2}$$

Two properties of the model are evident in the dynamic equation (eq. [D2]). First, the behavior of the signaler is driven by behavioral differences between the two types of listener—a type of decision rule that simply cannot occur with a single kind of listener. Second, the opposing listener, when faced with a signaler biased toward another listener, is selected to pursue a matching strategy. Whether the honesty of signals is high or not, the opposing listener (listener 1) should attempt to adjust its belief levels in the exact same way as the accomplice (listener 2). This is not a direct calculation by listener 1 but an indirect result of competing with the advantaged listener.

There are four sets of equilibria for the model: the eight vertices of the signaling cube, the back edge ($p = q = 0$), the front edge ($p = q = 1$), and the interior line defined by $t^* = 1/2$ and $p = q$. These are illustrated in figure D1. The isolated vertex equilibria are hyperbolic saddles.

Each equilibrium edge is partitioned by its intersection with the interior equilibrium line at $t = 1/2$, represented by the small circles in figures D1 and D2. One region is nonhyperbolic stable (green), while the other is nonhyperbolic and repelling (red). On the interior equilibrium line, the eigenvalue 0 has multiplicity three, so near the line there is a slowing of the evolutionary rate. The only corresponding eigenvector is $\mathbf{i} + \mathbf{j}$.

For the results shown in figure D2, we took $W = \epsilon = 1$. Examine the top right portion of the upper left diagram (fig. D2*a*), which represents the dynamics when listener 1's belief is held at 0 (face *F*4). In this situation, the signaler always has an incentive to tell the truth because his partner always has equal or greater belief. The evolution of the system demonstrates, however, that the signal system can still destabilize, depending on the initial conditions. There is a "belief terminus" on the face that is the separatrix between the basins of attraction in the face for the lower portion of edge $p = q = 0$ and the vertex (0, 1, 1).

The opposing face $p \equiv 1$, represented in the front left area of figure D2*b*, shows a similar response when the opponent listener always believes. In this case, there is still a belief terminus or separatrix between signal collapse and stability attractors. The system can still rebound or return to a state of belief, even if the preferred listener is initially (and perpetually) not performing as well as his competitor. These separatrices are actually the intersections between a larger boundary surface and the two faces. The entire boundary surface is represented in figure D2*d*. Note that the interior equilibrium line is contained within this surface. Trajectories whose initial condition lies in the boundary surface asymptotically converge to the interior equilibrium line.

Despite being in one basin or another, the transient behavior of variables need not be monotonic. The interior trajectories shown in figure D2*c* display considerable twisting before finally terminating at the edge attractors.

The dynamics of the two-listener interaction contrast markedly with the results for a single type of listener. If the signaler is communicating with his accomplice, listener 2, the payoff matrix (eq. [C1]) is

$q = 1$	W, ϵ	$-W, 0$
$q = 0$	$-W, 0$	W, ϵ
	$t = 1$	$t = 0$

(D3)

This game satisfies equation (C4) but not equation (C5)—the listener and signaler are in cahoots, so the signaler does not gain by deceiving the listener. In this game $c_1 - c_2 = \epsilon > 0$, and $c_3 - c_4 = -\epsilon < 0$. Consequently, (0, 0) and (1, 1) are stable equilibria, and there is an unstable interior saddle at (0.5, 0.5). The stable manifold of the saddle serves as the separatrix between the basins of attraction for the vertex (1, 1) and the breakdown vertex (0, 0). However, at either equilibrium the signal is honest, in the sense that signaler and accomplice both get the maximum possible reward because the listener guesses correctly whether the signaler is telling the truth. From the listener's perspective, the signal is always a reliable indicator of the true state of nature.

In contrast, if the signaler is engaged with only the opposing listener 1, then the payoff matrix (eq. [C1]) becomes

$p = 1$	$W, 0$	$-W, \epsilon$
$p = 0$	$-W, \epsilon$	$W, 0$
	$t = 1$	$t = 0$

(D4)

These payoffs satisfy equations (C4) and (C5), and we have $c_3 - c_4 = \epsilon > 0$ so we get cycling as in figure C1. Because of the cycling, the listener cannot reliably use the signal as an indicator of either nature state 1 or nature state 2. Thus, the full three-player game produces a potential result not seen in either two-player game, a persistent intermediate level of dishonesty.

APPENDIX E

Ambush Game with One Type of Listener

With only females as listeners in the ambush game, the two-player payoff matrix (eq. [C1]) is

	Attract mate ($t = 1$)	Ambush ($t = 0$)
Enter ($p = 1$)	$M, M - E$	$-A, 0$
Avoid ($p = 0$)	$B_2, -E$	$B_2, 0$

(E1)

The parameter combinations defining the edge flows are

$$\begin{aligned} a_1 - a_3 &= M - B_2 > 0, \\ a_2 - a_4 &= -(A + B_2) < 0, \\ c_1 - c_2 &= M - E > 0, \\ c_3 - c_4 &= -E \leq 0. \end{aligned} \tag{E2}$$

As in the full game the outcome depends on whether E is positive or 0.

1. If $E > 0$, then the edge flows suggest bistability of (0, 0) and (1, 1), and this can be proved to hold in equation (C2); if $t(0)$ and $p(0)$ are both sufficiently large, then $t' > 0$ and $p' > 0$, implying convergence to (1, 1), and, similarly, there is convergence to (0, 0) if $t(0)$ and $p(0)$ are small. The conditions for an interior equilibrium to exist are satisfied, but $\alpha\gamma = (c_1 - c_2)(a_1 - a_3) > 0$, so the interior equilibrium is a saddle. All trajectories therefore converge to (0, 0) or (1, 1), except for the saddle and its stable manifold.

2. If $E = 0$, we have the nongeneric situation where one of the coefficients in equation (C2) is exactly 0. But unless $p = 0$ (meaning that all females always ignore a caller), it pays for a caller to give an honest signal; hence $t \to 1$ and therefore $p \to 1$ because $a_1 - a_3 > 0$.

So in either case, the long-run behavior is that a calling male does not engage in ambushing if only females enter the territory. With $E = 0$, the end result is honest calls that females trust. With $E > 0$, if the ambushing rate is initially high, females will avoid entering the territory, and callers will therefore choose to avoid the cost of being ready to mate—the signal has then inverted its meaning and is an honest warning of intent to ambush, which females respond to appropriately by looking for better chances elsewhere. Otherwise the parties converge again to (1, 1). So in all cases, the caller-female game leads to an honest and correctly interpreted signal.

If satellite males are the only listeners, the only possible outcome is convergence to (0, 0). If a satellite enters, it might get ambushed, but it cannot get a payoff because there are no females. Therefore, satellites should always stay out. The best option for callers is then to give the cost-free deceptive signal. So again, the signal is inverted into an accurate warning of intent to ambush.

APPENDIX F

Territory Challenge Game with One Type of Listener

If territory holders face only big challengers, the reduced payoff matrix is

	True signal	False signal
Respect	B_1, T	$B_1, T + \epsilon$
Challenge	$\frac{T}{2} - F_1 - I, \frac{T}{2} - F_1 + I$	$T - F_3, -F_3$

(F1)

The parameter combinations defining the edge flows are

$$\begin{aligned} a_1 - a_3 &= B_1 - \left(\frac{T}{2} - F_1 - I\right) > 0, \\ a_2 - a_4 &= B_1 - T - F_3 < 0, \\ c_1 - c_2 &= -\epsilon \le 0, \\ c_3 - c_4 &= \frac{T}{2} - F_1 + I + F_3 > 0. \end{aligned} \tag{F2}$$

So if there is a territory bonus for small individuals, we have cycling as in figure C1. At any time there will be both big and small signalers and a mix of strategies among potential challengers. However, if the bonus for small territory holders is absent, then all points on the complete-belief edge $p = 1$ are neutrally stable equilibria. The long-term behavior is convergence onto this edge at some point, which depends on the initial state. In that case listeners always believe, despite the presence of many deceptive (small) signalers.

If territory holders face only small challengers, the reduced payoff matrix is

	True signal	False signal
Respect	B_2, T	$B_2, T + \epsilon$
Challenge	$-F_3, T - F_3$	$\frac{T+\epsilon}{2} - F_2 - I, \frac{T+\epsilon}{2} - F_2 + I$

(F3)

The parameter combinations defining the edge flows are

$$\begin{aligned} a_1 - a_3 &= B_2 + F_3 > 0, \\ a_2 - a_4 &= B_2 - \left(\frac{T + \epsilon}{2} - F_2 - I\right), \\ c_1 - c_2 &= -\epsilon \le 0, \\ c_3 - c_4 &= (T - F_3) - \left(\frac{T + \epsilon}{2} - F_2 + I\right). \end{aligned} \tag{F4}$$

Because the directions of the edge flows are not all determined by our parameter assumptions, there are several different possibilities.

1. Do big or small defenders get a higher payoff when challenged by a small individual? This determines the sign of $c_3 - c_4$.

2. Is a small challenger's background opportunity larger than his expected payoff from challenging a small defender? This determines the sign of $a_2 - a_4$.

Suppose first that $\epsilon > 0$. If $a_2 - a_4 > 0$, then the system approaches the vertex (0, 1) in (t, q) space; challenges never occur (because of the high background opportunity), and small signalers become predominant because of their

greater benefit from territory holding. If instead $a_2 - a_4 < 0$ and also $c_3 - c_4 < 0$, then the asymptotic equilibrium is complete breakdown (0, 0); signals are challenged, and small signalers predominate because of their higher payoff against small challengers. Finally, if $a_2 - a_4 < 0$ and $c_3 - c_4 > 0$, trajectories oscillate as in figure C1.

Without a territory bonus to small defenders, the behavior observed changes slightly. In cases where the asymptotic attractor was successful in deception ($t = 0$, $q = 1$) for $\epsilon > 0$, the corresponding attractor is the entire complete-belief edge ($q = 1$, $0 \le t \le 1$). Cycling is replaced with attraction to the higher t region of the complete-belief edge $q = 1$. Finally, in situations leading to breakdown with $\epsilon > 0$, that equilibrium is coupled with the complete belief attractor $q = 1$, $0 \le t \le 1$.

Literature Cited

Abrams, P. A. 1999. The adaptive dynamics of consumer choice. American Naturalist 153:83–97.

———. 2005. "Adaptive dynamics" vs. "adaptive dynamics." Journal of Evolutionary Biology 18:1162–1165.

Abrams, P. A., Y. Harada, and H. Matsuda. 1993. On the relationship between quantitative genetics and ESS models. Evolution 47:982–985.

Adams, E. S., and R. L. Caldwell. 1990. Deceptive communication in asymmetric fights of the stomatopod crustacean *Gonodactylus bredini*. Animal Behaviour 39:706–716.

Adams, E. S., and M. Mesterton-Gibbons. 1995. The cost of threat displays and the stability of deceptive communication. Journal of Theoretical Biology 175:405–421.

Alcock, J., and A. Forsyth. 1988. Post-copulatory aggression toward their mates by males of the rove beetle *Leistotrophus versicolor* (Coleoptera: Staphylinidae). Behavioral Ecology and Sociobiology 22:303–308.

Arak, A. 1988. Callers and satellites in the natterjack toad: evolutionarily stable decision rules. Animal Behaviour 36:416–432.

Arnold, S. J. 1978. The evolution of a special class of modifiable behaviors in relation to environmental pattern. American Naturalist 112:415–428.

Backwell, P. R. Y., J. H. Christy, S. R. Telford, M. D. Jennions, and N. I. Passmore. 2000. Dishonest signalling in a fiddler crab. Proceedings of the Royal Society B: Biological Sciences 267:719–724.

Bee, M. A., S. A. Perrill, and P. C. Owen. 2000. Male green frogs lower the pitch of acoustic signals in defense of territories: a possible dishonest signal of size? Behavioral Ecology 11:169–177.

Bradbury, J. W., and S. L. Vehrencamp. 1998. Principles of animal communication. Sinauer, Sunderland, MA.

Byrne, R. W., and N. Corp. 2004. Neocortex size predicts deception rate in primates. Proceedings of the Royal Society B: Biological Sciences 271:1693–1699.

Clement, T. S., K. E. Grens, and R. D. Fernald. 2005. Female affiliative preference depends on reproductive state in the African cichlid fish, *Astatotilapia burtoni*. Behavioral Ecology 16:83–88.

Cohen, Y., T. L. Vincent, and J. S. Brown. 1999. A *G*-function approach to fitness minima, fitness maxima, ESS and adaptive landscapes. Evolutionary Ecology Research 1:923–942.

Dawkins, M. S., and T. C. Guilford. 1991. The corruption of honest signalling. Animal Behaviour 41:865–873.

Doutrelant, C., and P. K. McGregor. 2000. Eavesdropping and mate choice in female fighting fish. Behaviour 137:1655–1669.

Doutrelant, C., P. K. McGregor, and R. F. Oliveira. 2001. The effect of an audience on intrasexual communications in male Siamese fighting fish, *Betta splendens*. Behavioral Ecololgy 12:283–286.

Earley, R., and L. A. Dugatkin. 2002. Eavesdropping on visual cues in green swordtails (*Xiphoporus helleri*): a case for networking. Proceedings of the Royal Society B: Biological Sciences 269:943–952.

Gardner, R., and M. R. Morris. 1989. The evolution of bluffing in animal contests: an ESS approach. Journal of Theoretical Biology 137:235–243.

Getty, T. 1997. Deception: the correct path to enlightenment? Trends in Ecology & Evolution 12:159–160.

Godfray, H. J. C. 1991. Signaling of need between parents and offspring. Nature 352:328–330.

———. 1995. Signaling of need between parents and young: parent-offspring conflict and sibling rivalry. American Naturalist 146:1–24.

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

Hofbauer, J., and K. Sigmund. 1988. Evolutionary games and replicator dynamics. Cambridge University Press, Cambridge.

———. 2003. Evolutionary game dynamics. Bulletin of the American Mathematical Society 40:479–519.

Hughes, M. 2000. Deception with honest signals: signal residuals and signal function in snapping shrimp. Behavioral Ecology 11:614–623.

Johnstone, R. A. 1994. Honest signalling, perceptual error and the evolution of "all-or-nothing" displays. Proceedings of the Royal Society B: Biological Sciences 256:169–175.

———. 1995. Honest advertisement of multiple qualities using multiple signals. Journal of Theoretical Biology 177:87–94.

———. 1998. Game theory and communication. Pages 94–117 *in* L. A. Dugatkin and H. K. Reeve, eds. Game theory and animal behavior. Oxford University Press, New York.

Johnstone, R. A., and A. Grafen. 1992. The continuous Sir Philip Sidney game: a simple model of biological signalling. Journal of Theoretical Biology 156:215–234.

———. 1993. Dishonesty and the handicap principle. Animal Behaviour 46:759–764.

Kokko, H. 1997. Evolutionarily stable strategies of age-dependent sexual advertisement. Behavioral Ecology and Sociobiology 41:99–107.

Lucas, J. R., R. D. Howard, and J. G. Palmer. 1996. Callers and satellites: chorus behaviour in anurans as a stochastic dynamic game. Animal Behaviour 51:501–518.

Maynard Smith, J. 1974. The theory of games and the evolution of animal conflicts. Journal of Theoretical Biology 47:209–221.

———. 1979. Game theory and the evolution of behaviour. Proceedings of the Royal Society B: Biological Sciences 205:475–488.

———. 1991. Honest signalling: the Philip Sidney Game. Animal Behaviour 42:1034–1035.

———. 1994. Must reliable signals always be costly? Animal Behaviour 47:1115–1120.

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

Maynard Smith, J., and G. A. Parker. 1976. The logic of asymmetric contests. Animal Behaviour 24:159–175.

Maynard Smith, J., and G. R. Price. 1973. The logic of animal conflict. Nature 246:15–18.

McGregor, P. K. 1993. Signalling in territorial systems: a context for individual identification, ranging and eavesdropping. Philosophical Transactions of the Royal Society B: Biological Sciences 340: 237–244.

McGregor, P. K., and T. Dabelsteen. 1996. Communication networks. Pages 409–425 *in* D. E. Kroodsma and E. H. Miller, eds. Ecology and evolution of acoustic communication in birds. Cornell University Press, Ithaca, NY.

McGregor, P. K., and T. Peake. 2000. Communication networks: social environments for receiving and signalling behaviour. Acta Ethologica 2:71–81.

McLain, D. K., and A. E. Pratt. 1999. The cost of sexual coercion and heterospecific sexual harassment on the fecundity of a host-specific, seed-eating insect *Neacoryphus bicrucis.* Behavioral Ecology and Sociobiology 46:164–170.

Munn, C. A. 1986. Birds that cry wolf. Nature 319:143–145.

Naguib, M., and D. Todt. 1997. Effects of dyadic interactions on other conspecific receivers in nightingales. Animal Behaviour 54: 1535–1543.

Nowak, M. A., and K. Sigmund. 2004. Evolutionary dynamics of biological games. Science 303:793–798.

Oliveira, R. F., P. K. McGregor, and C. Latruffe. 1998 Know thine enemy: fighting fish gather information from observing conspecific interactions. Proceedings of the Royal Society B: Biological Sciences 265:1045–1049.

Otter, K., P. K. McGregor, A. M. R. Terry, F. R. L. Burford, T. M. Peake, and T. Dabelsteen. 1999. Do female great tits (*Parus major*) assess males by eavesdropping? a field study using interactive song playback. Proceedings of the Royal Society B: Biological Sciences 266:1305–1310.

Reed, C., T. G. O'Brien, and M. F. Kinnaird. 1997. Male social behavior and dominance hierarchy in the Sulawesi crested black macaque *Macaca nigra.* International Journal of Primatology 18: 247–260.

Rowell, J. T. 2003. Diffusive food webs and signaling dynamics of populations. PhD diss. Cornell University, Ithaca, NY.

Searcy, W. A., and S. Nowicki. 2005. The evolution of animal communication: reliability and deception in signaling systems. Princeton University Press, Princeton, NJ.

Steger, R., and R. L. Caldwell. 1983. Intraspecific deception by bluffing: a defense strategy of newly molted stomatopods (Arthropoda: Crustacea). Science 221:558–560.

Tomkins, J. L., and L. W. Simmons. 2002. Measuring relative investment: a case study of testes investment in species with alternative male reproductive tactics. Animal Behaviour 63:1009–1016.

Wiley, R. H. 1983. The evolution of communication: information and manipulation. Pages 156–189 *in* T. R. Haliday and P. J. B. Slater, eds. Animal behavior. Vol. 2. Communication. Blackwell Scientific, Oxford.

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

———. 1977*a.* The cost of honesty (further remarks on the handicap principle). Journal of Theoretical Biology 67:603–605.

———. 1977*b.* Reliability in communication systems and the evolution of altruism. Pages 253–259 *in* B. Stonehouse and C. Perrins, eds. Evolutionary ecology. Macmillan, London.

———. 1981. Natural selection, sexual selection and the selection of signals. Pages 133–138 *in* G. G. E. Scudder and J. L. Reveal, eds. Evolution today. Proceedings of the Second International Congress of Systematics and Evolution. Carnegie-Mellon University, Pittsburgh, PA.

———. 1987. The theory of signal selection and some of its implications. Pages 305–327 *in* V. P. Delfino, ed. International symposium of biological evolution. Adriatica Editrice, Bari.

Associate Editor: Peter D. Taylor
Editor: Michael C. Whitlock