Errors, Exaggeration, and Deception in Animal Communication

R. Haven Wiley

MY CONTENTION in this chapter is that animals make mistakes and that this simple but neglected circumstance has deep implications for the evolution of animal communication. The possibility of error is explicitly incorporated into mathematical theories of communication and choice, in particular, information theory, signal detection theory, and decision theory. Although the basic concepts of these theories are neither new nor complex, they have yet to be integrated into evolutionary theories of animal communication. An objective of this chapter is to show, by expanding themes introduced earlier (Wiley 1983), that these basic concepts provide explanations for the evolution of some fundamental features of communication. In particular, they can provide sufficient conditions for the evolution of exaggerated displays and for the evolutionary stability of deception. This approach also indicates that some of the parameters needed to understand the evolution of communication have been overlooked.

THREE EXAMPLES OF SITUATIONS IN WHICH ERROR OCCURS

To begin, consider three situations in which animals might make mistakes in responding to signals. In the present context, a mistake or error is an evolutionarily inappropriate response, one that reduces chances for the spread of genes associated with the response. Some definitions of terms used in the study of communication warrant attention in a later section. For the moment, an evolutionary perspective provides an operational definition of error. The following examples illustrate the possibilities for error during communication at long range, in dense aggregations, and in the presence of deception.

Long-Range Communication with Song

First, as an example of communication at long range, consider a male territorial passerine bird listening for conspecifics' songs in a forest. Appropriate responses to those songs might include approaching any conspecific male that sings within the listener's territory or countersinging with a neighbor. Song at close range and at full power might almost always evoke an appropriate response; listeners, in other words, would make few, if any, errors. Yet even slight hesitation by a temporarily distracted listener or confusion of songs with superficially similar sounds in the environment could make the listener's response less than optimal. At a distance, the chances for error increase. The signal heard by the listener is attenuated and
distorted, sometimes to the extent that detection fails completely. Possibilities for error increase further if the listener’s task is more complicated than simply detecting conspecifics; for instance, it might include recognizing different neighbors’ songs, localizing the source, discriminating different song types, or associating songs with memories of previous locations or interactions. In addition, there is the possibility that singers might deceive or manipulate listening rivals by minimizing possibilities for recognition or localization (Krebs 1977; Morton 1986). Then of course the possibilities for error increase still more.

Current information suggests that communication between territorial birds by means of song requires complex discriminations. Territorial male passerines discriminate between conspecific songs and others, between neighbors’ and strangers’ songs, and between songs of individual neighbors in normal and abnormal locations (Falls and Brooks 1975; Wiley and Wiley 1977; Becker 1982; Falls 1982; Godard 1991, 1992; Stoddard et al. 1991). When individuals have repertoires of distinct song types, their use of these song types differs, at least in some species, with location within their territories or with probabilities of subsequent actions (Lein 1978; Smith et al. 1977; Schroeder and Wiley 1983; Temeiner 1986; Dabelsteen and Pedersen 1990). In a few cases, experimental playbacks of songs have shown that males respond differently to these variants (Järvi et al. 1980; Schroeder and Wiley 1983; Dabelsteen and Pedersen 1990). Females also recognize conspecific and even individual males’ songs (King and West 1983; Searcy et al. 1981; Searcy 1990; Wiley et al. 1991). Woodland birds can also judge the distance to a singer by attenuation or degradation of songs (Richards 1981b; MacGregor and Krebs 1984; MacGregor 1991). A listening bird’s response thus depends on a set of discriminations based on the features of a song. Communication with song often occurs over tens to hundreds of meters and in the presence of other species with some similar features in their songs, conditions that make perfect detection and recognition of songs unlikely.

Experimental playbacks of recorded songs in the field elicit variable responses, often frustratingly so. In analyzing these responses, investigators have always focused on the group means across subjects, rather than on variation in responses by individual subjects to replicate presentations. The variation, however, emphasizes that inconsistency in responses persists even to a standardized signal in standardized situations. An understanding of error in communication requires an analysis of this variation.

Communication in Mating Aggregations of Frogs

As a second example of the possibility for error in communication, consider frogs that mate in aggregations. This situation raises the possibility of error in females’ choice of mates. In an experiment designed to determine the effects of background sounds on acoustic communication, Gerhardt and Klump (1988a) tested the ability of female green treefrogs (Hyla cinerea) to orient toward a male’s calls when they were masked by the sounds of a large chorus. The female’s problem was to detect and locate an intermittent call in the presence of continuous background sounds of similar frequency. Only when the intensity of a male’s call equaled or exceeded that of the background sound did most females succeed. As a result of this limitation, a female presumably detects only some three to five males at a time in a large chorus.

In mixed choruses, there is also the possibility of interference from calls of different species. Frogs active at the same season and in the same habitats sometimes avoid overlap in calling (Littlejohn and Martin 1969; Zelick and Narins 1982;
Schwartz and Wells 1983a, 1983b; Schwartz 1987) and differ distinctly in the spectral or temporal patterning of their calls (Littlejohn 1977). Nevertheless, the partitioning
is sometimes incomplete. For example, *Hyla ebraccata* often calls within choruses of
*H. microcephala* in Panama; the spectrum of *ebraccata*’s calls broadly overlaps the
lower peak in the spectrum of its congener’s calls. To reduce masking, *ebraccata* often
remain silent during cyclic peaks of calling by the louder and denser *microcephala*
(Schwartz and Wells 1983a, 1983b). In *H. microcephala* and *H. versicolor*, playback
experiments suggest that both males and females are more likely to respond appro-
priately to nonoverlapping calls (Schwartz 1987). The possibility of inappropriate
responses, including failure to respond, thus seems plausible in the presence of
interfering calls. The possibility for errors might depend on the distance of communi-
cation, although Gerhardt and Klump (1988b) showed that female barking treefrogs
(*H. gratiosa*) preferred sounds of choruses that contained conspecific males over
those that did not at a distance of 160 m. Experiments with green treefrogs suggest
that low-frequency components in the males’ calls might attract females from a
greater distance than the high-frequency components (Gerhardt 1987).

Females approaching these mixed choruses in some cases must make fine dis-
criminations. *H. chrysoceleis* and *H. versicolor* differ primarily in the temperature-
dependent pulse rates of their calls. Female *versicolor* prefer lower pulse rates than
do female *chrysoceleis* at any temperature. They can discriminate pulse rates differing
by much less than a factor of two. In addition, when pulse rates differ by at least a
factor of two, females’ preferences are stable even when the sound pressure level
of the nonpreferred call exceeds that of the preferred one by 18 dB (Gerhardt 1982).
Even though *chrysoceleis* is diploid and *versicolor* tetraploid, there appears to be some
hybridization where the two species occur together (Gerhardt pers. comm.), so
errors in choice of mates apparently do occur. Errors in mating might occur even
more often than indicated by hybridization.

Hybridization between sympatric *H. cinerea* and *H. gratiosa* has been carefully
documented (Gerhardt et al. 1980). In addition, *H. squirella* also shares breeding sites
with these two species. When given no choice, females of each species occasionally
approach playbacks of one of the other species’ calls (Oldham and Gerhardt 1975),
even though some features, including the peak frequency in the lower of two spec-
tral bands in the calls of each species, have little or no overlap. In two-choice tests,
female *cinerea* prefer synthetic calls with a low-frequency peak (LFP) of 900 Hz over
those with an LFP lower than or higher than 800–1,000 Hz and thus discriminate
against sounds with LFPs typical of *gratiosa* and *squirella*, respectively. Nevertheless,
some female *cinerea* chose the alternative calls. Furthermore, discrimination is less
clear when females have a simultaneous choice of four calls (Gerhardt 1982, 1987).
In general, reducing the difference in frequency between alternative calls and in-
creasing the number of simultaneous choices both reduce a female’s selectivity (fig.
7.1). Furthermore, as Gerhardt (1982) indicates, females can hardly rely on experi-
ence in making these discriminations, in the absence of extended associations of
parents and offspring or of mates. Despite these possibilities for errors by females,
hybrids are rarely found (Schlefer et al. 1986). However, when heterospecific mat-
ings produce less viable offspring, the frequency of hybrids underestimates the
number of mismatings.

Intraspecific Deception

A third source of error in animal communication is deception. Deception occurs
when a signaler, in specified circumstances, gains from a receiver’s response to a
Fig. 7.1 Female green tree frogs are less discriminating in four-choice than in two-choice experiments. (A) When females are given a choice between two speakers, one playing a naturalistic synthetic call with a low-frequency peak (LFP) at 900 Hz and another playing a call at one of four other frequencies, they never respond to calls with LFPs at 600 or 1,200 Hz \((n \geq 8\) females for each bar). (B) When presented with four choices of synthetic calls, females sometimes find their way to the speakers presenting calls with LFPs at 600 or 1,200 Hz \((n = 60\) females). In both experiments sounds were presented to females at 75 dB. (Adapted from Gerhardt 1982.)

signal while the receiver loses. This situation arises when a signaler takes advantage of a receiver’s rules for decoding signals (Wiley 1983). Gains and losses in this context are changes in the signaler’s and receiver’s expected survival and reproduction that in turn affect the spread of genes associated with signaling and responding. The possibility of this sort of interaction between signalers and receivers was first emphasized by Dawkins and Krebs (1978) in their discussion of manipulation in communication.

Many recently documented cases of intraspecific deception fall into two categories: mimicry of females by males to gain surreptitious matings, and use of alarm calls to gain temporary advantages in competition for food or territories. An example of the former is the occurrence of small, female-like males in several species of fish and crustaceans. In small populations of the bluehead wrasse \((Thalassoma bifasciatum)\), a Caribbean reef fish, most females mate at preferred sites monopolized by large, terminal-phase males, which have completed protogynous sex reversal. Small males form spawning groups or attempt to accompany females spawning with large males (Warner and Hoffman 1980; Warner 1984, 1987). Either way they presumably fertilize a small proportion of any one spawning female’s eggs. Large males aggressively exclude small males from their territories. The small males’ success in parasitizing large males depends on their close resemblance to females, which makes discrimination difficult for terminal-phase males. In large populations, terminal-phase males cannot monopolize the preferred mating sites, and thus females more often spawn with small males. As expected from their relative success in mating, small
males constitute a greater proportion of large populations. Because individuals do not change reefs after settling, small males facultatively allocate their resources to either growth or spawning in accordance with the size of the reef they inhabit and their consequent chance for reproduction at an early age (Warner 1984).

In other cases, individuals can change tactics within their lifetimes. In the scorpionfly Hylobittacus apicalis, males can alternate between hunting for prey to use as nuptial offerings to attract females—behavior that risks capture in spiders’ webs—and posing as females to steal other males’ offerings (Thornhill 1979). A male with a nuptial offering sometimes recognizes a deceptive male in time to protect its offering, but on other occasions presents it to the deceptive male and thus loses it. In both wrasses and scorpionflies, males with an inherent advantage in mating confront the problem of discriminating between the similar appearances of females and deceptive males. Errors, either responding to a deceptive male as if it were a female or to a female as if it were a deceptive male, result in lower success in mating.

Examples of deceptive use of alarm calls include the great tits (Parus major) studied by Möller (1988). Tits at an artificial feeding station often flee when they hear the distinctive alarm call usually uttered by another tit after spotting a hawk. Occasionally a tit prevented from feeding by a dominant opponent utters an alarm call and thus gains access to the feeding site after the opponent departs. Tits use this ploy only when a feeding site is occupied by a dominant opponent. In another example, individual antshrikes (Thamnomanes schistogynus) in mixed-species flocks in Amazonian forests produce a conspicuous call both when a raptor is spotted and when a nearby individual of some other species is about to seize a large insect. In the latter case, during the moment of hesitation by the nearby individual when it hears the possible alarm, the antshrike often catches the prey (Munn 1986a, 1986b). Spectrograms of calls produced in the two contexts do not differ. In these cases, a signal virtually indistinguishable from one used in other contexts serves to evoke a response inappropriate for the receiver. From the receiver’s point of view, failure to discriminate any subtle differences between the normal and the deceptive signals or their contexts results in an error.

**FUNDAMENTAL PROPERTIES OF A COMMUNICATION SYSTEM**

Before proceeding, it is worthwhile to review some fundamental features of communication and some basic definitions. A framework for studying errors in communication develops naturally from these basic concepts.

Definitions of deception or of errors in communication that rely entirely on evolutionary criteria will not suit everyone. Cognitive ethologists have tended to insist that true deception requires intention, an awareness by the signaler of its influence on the receiver’s state of mind (Woodruff and Premack 1979; Ristau 1991; Ristau and Robbins 1982; Cheney and Seyfarth 1990). The issue of intentions also surfaces in definitions of communication itself or, at least, of the distinctive human form of communication by means of language (Grice 1969; Austin 1973; Dennett 1987, 1988). One motivation for stipulating intentions as a criterion for communication has been to separate communication from behavior that is unconscious, conditioned, innate, or even accidental.

Any attempt to include intentions in the necessary conditions for communication faces the difficulty of specifying generally acceptable definitions for mental states in other individuals. It is not yet clear that it is possible to devise such defini-
tions for mental states like intentions (see, for instance, Putnam 1967; Bennett 1976; Harré 1984; Harris 1984; Dennett 1988; Cheney and Seyfarth 1990). In any case, the present evolutionary analysis does not require them. Nor does the present analysis require specification of the proximate behavioral or neural mechanisms of communication, but it assumes that they exist and can evolve.

Another frequent stipulation in definitions of communication, at least in the past, has been mutually advantageous interaction between sender and receiver. This condition seemed reasonable because it excludes such intuitively noncommunicative interactions as predation. However, it also excludes deception, as defined above, and indeed any possibility for error. It thus excludes some of the most interesting possibilities for communication.

Communication can be defined in a way that avoids these difficulties as any alteration in a receiver produced by a signaler by means of a signal. A signal is any pattern of energy or matter produced by one individual (the signaler) and altering some property of another (the receiver) without providing the power to produce the entire response. The last phrase excludes all interactions between individuals in which the actor overpowers the recipient; for instance, it makes a distinction between communication and predation. It sets aside all questions about the mechanisms involved, including intentions or other mental states, however they might be defined by behavioral or neural events. Indeed, these basic definitions apply to machines as well as animals.

In this approach, the components of any system of communication consists of a signaler, a signal, and a receiver. In this triplet of components, first introduced by Shannon and Weaver (1949) in their analysis of rates of information transfer, the signaler and receiver are unspecified mechanisms with outputs that are associated by means of a signal produced by one and detected by the other. This scheme differs crucially from the earlier triplet of signal, receiver, and referent. This latter triplet, diagrammed for the first time by Ogden and Richards (1923), emphasizes the relationship between a signal and its referent, or as Saussure (1959 [1915]) put it, between a "sound-image" and an associated "concept." This relationship, mediated by the receiver and often called interpretation or representation, raises all the issues about the operational definitions of mental states mentioned above.

Any scientific investigation of communication actually requires a fourth essential component, a nonparticipant observer, first introduced by Shannon and Weaver (1949) and emphasized by Cherry (1957). Of course, few observers can completely avoid all participation or interference in their subjects' interactions, so perhaps "minimally participant," or "scientific," observer is a more accurate term. It behooves a scientific observer to assess all possible influences on the subjects of study. Such an observer should be able to intercept the signal, often at different points in its propagation between signaler and receiver, and to measure behavioral and physiological events and states in both signaler and receiver.

In this view of communication, a receiver is an unspecified mechanism for associating signals and responses. The stipulation that the signal cannot produce the "entire response" is important, because of course the signal must provide enough energy to effect some response in the receiver's sensors; yet some, usually most, of the energy to produce the response must come from the receiver itself. A receiver must thus include, in some form, both transducers and amplifiers (or sensors and effectors). As a behavioral, neural, and, ultimately, physical mechanism, it must also follow specifiable rules; alterations in the receiver as a result of a signal must be a function of the signal and the current state of the receiver (Wiley 1983). An
alteration in a receiver might include overt changes in effectors, such as a movement of limbs, or vocal cords, that could provide a signal for another receiver. It might also go unnoticed by an observer and thus remain covert. The alterations in the receiver of a signal might involve great complexity.

These possibilities allow us to decompose any act of communication into components each one of which admits further analysis into smaller, but conceptually similar, components. Thus one analysis might focus on a signal, receiver, and response that consisted of the mating calls of a particular species of frog, conspecific females, and movement toward a speaker. Another might focus on artificially synthesized sounds, auditory receptors in the basilar papillae of a frog, and impulses in the eighth cranial nerve; another on sounds, perhaps unspecified neural circuits, and evoked potentials in the tectum. Responses might include complex patterns of neural activity without unique motor consequences and not measurable with currently available equipment.

These examples emphasize that many responses are apparent only to a scientific observer with special apparatus and that some remain covert even to observers with the best available equipment. Note also that any signal-receiver-response system consists of a nested hierarchy of conceptually similar systems and that a hierarchy of these systems constitutes the full description of any act of communication. An essential feature of this view of communication is the emphasis on associations between signals and responses. A receiver at any level of analysis, viewed as a whole, is a mechanism that associates signals with responses.

The simplest way to think of such a system is to consider two alternative (mutually exclusive) signals associated by a receiver with two alternative responses. Thus incoming signals are of two types, to each of which the receiver makes one of two kinds of responses. More complex situations are elaborations of this simple one. For instance, to study detection of a single type of signal, we can divide stimulation impinging on the receiver into two categories, signal present or signal absent. To study recognition of alternative signals, we can compare two signals, such as conspecific and heterospecific songs or familiar and unfamiliar ones. We can divide responses into two categories, such as approaching or not, vocalizing for longer or shorter periods, or, in the case of humans, saying "yes" or "no."

A scientific observer of this simplest system can identify four possible stimulus-response pairs, each of two signals crossed with each of two responses. The observer must accurately differentiate the two categories of signals and the two categories of responses. It is this situation that signal detection theory analyzes, although the mathematical elaboration of the theory requires additional assumptions: that the two stimulus states differ along a single dimension (intensity, for instance); and that both include random perturbations with Gaussian distributions and equal variances along this dimension (Green and Swets 1966; Egan 1975). These assumptions of unidimensionality and homoscedasticity are not essential for some basic conclusions about communication, however.

In cases in which one of the two categories of response is appropriate for one of the two categories of signal, the four signal-response pairs have clear interpretations (table 7.1). When the appropriate signal is present, responses are either correct detections (CD) or missed detections (MD), and, when the appropriate signal is absent, responses are either false alarms (FA) or correct rejections (CR). These four possibilities arise whenever two signal categories are not perfectly discriminated by the receiver. Unpredictable variation in the properties of signals, from whatever source, can produce overlap between two signal categories, as sensed by a receiver, so the receiver
### Table 7.1 Four Possibilities when a Signal Is Not Perfectly Discriminated by a Receiver

<table>
<thead>
<tr>
<th>Signal</th>
<th>Response Appropriate for Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present</td>
<td>Present: Correct detection</td>
</tr>
<tr>
<td>Absent</td>
<td>Absent: Missed detection</td>
</tr>
<tr>
<td>Absent</td>
<td>Absent: Correct rejection</td>
</tr>
<tr>
<td>Present</td>
<td>Present: False alarm</td>
</tr>
</tbody>
</table>

cannot perfectly distinguish them. Note that a scientific observer, by means of a special vantage or equipment, might be able to distinguish two signal categories that a receiving animal could not. Whenever a receiver cannot perfectly distinguish two signal categories, it cannot simultaneously reduce both its probability of false alarm and its probability of missed detection to zero.

The fundamental consequence of this situation is that a receiver cannot simultaneously maximize its probability of correct detection and minimize its probability of false alarm. This conclusion does not depend on any assumptions concerning the distribution of perturbations or equality of variances. It applies whenever the categories of signals (including signal and no signal) preclude perfect discrimination by the receiver’s sensors for any reason.

## ERROR AND NOISE IN ANIMAL COMMUNICATION

Any conclusion about "error" as an evolutionarily inappropriate response by a receiver requires estimates of the spread of genes associated with responses, as indicated, for instance, by the receiver’s (and its kin’s) probabilities of survival and reproduction. It is useful to distinguish error, in this sense, from noise.

The term "noise," in everyday usage, comes close to meaning irrelevant masking energy. Quite a different view emerges from Shannon and Weaver’s (1949, 20-21, 66-70) discussion of noise as it applies to information theory. They define error as equivocation, $H_y(x)$, the average uncertainty (or entropy) in the input from the source (x) when the output from the receiver (y) is known:

$$(1/n) \sum_x \sum_y p_{x|y} \log_2 p_{x|y}$$

where $p_{x|y}$ is the probability of the $x$th signal category provided a response of the $y$th response category has occurred and $n$ is the number of possible responses. In terms appropriate for behavioral interactions, equivocation is a scientific observer’s average uncertainty about which signal (x) has occurred when the observer knows only the response of a receiver (y). Transmission of information between signaler and receiver occurs at the maximal rate when every signal category produces a unique response category. The rate of transmission, $R$, then equals the uncertainty in either the signals or the responses considered alone: $R_{\text{max}} = H(x) = H(x) = H(y)$. To obtain the actual rate of transmission of information, an observer must subtract the measure of equivocation from the uncertainty in the signals alone:

$$R = H(x) - H_y(x).$$
Note that if a signal is completely predictable, so that \( H(x) = 0 \) and thus \( H_y(x) = 0 \), then \( R = 0 \). Only when signals occur with some uncertainty can they transfer information. In addition, if for any reason the signaler’s behavior has no association with the receiver’s, so that \( H_y(x) = H(x) \), then again \( R = 0 \). Thus \( R \) measures the amount of information that signals actually convey to a receiver. A mathematically equivalent statement,

\[
R = H(y) - H_x(y),
\]

expresses the rate of transmission as the difference between uncertainty in the responses and the average conditional uncertainty of responses when the signal is known. In the first equation, \( H_y(x) \) represents the amount of information needed to correct the received message. In the second equation, \( H_x(y) \) represents the portion of the uncertainty in the receiver’s responses that is “due to noise,” as Shannon and Weaver put it.

Therefore, for these authors, transmission is \textit{noiseless} when there is no uncertainty about the response once the signal is known. This perspective merges with our conventional view that noise is masking energy, so that responses are functions of both signals and noise. On the other hand, according to Shannon and Weaver, transmission is \textit{error-free}, or needs no correction, when there is no uncertainty about the signal when the response is known.

In applying these ideas to animals, some problems arise. As engineers, Shannon and Weaver view communication as a reconstruction by the receiver of symbols exactly as they originate from the source—a dot for a dot, a dash for a dash, or a letter for a letter. In animal communication, however, the action of the receiver is often different from the action of the signaler. We cannot judge errors by comparing the form of the output with that of the input. Instead, it makes more sense to define errors in terms of the evolutionary appropriateness of the response to a signal in a particular situation. Just as Shannon and Weaver indicate in their discussion of error, the inappropriateness of a response must be determined by an observer with knowledge of the signaler, signal, receiver, and context.

The engineer’s objective of reconstructing the signal from the response assumes that the signaler and receiver share the same interests in communication. In considering the evolution of animal communication, this assumption does not necessarily apply. However, the engineer’s definitions of noise and error do apply. From the signaler’s perspective, a signal should evoke a particular response from an intended receiver in a particular situation; any uncertainty about the response following a signal is noise from the signaler’s viewpoint. Thus, for a signaler, communication is noise-free when \( H_y(x) = 0 \). From a receiver’s perspective, a response should be made to a particular kind of signal in a particular situation; any uncertainty about the preceding signal raises the possibility of an inappropriate response, an error for the receiver. Thus, for a receiver, communication is error-free when \( H_y(x) = 0 \). These two conditional uncertainties are usually nonzero and not necessarily equal.

For a signaler, noise results from anything that reduces the probability that a signal will evoke a particular response from the intended receiver. For a male frog in a chorus, noise could result from other males’ calls, environmental attenuation and degradation of its own calls, or reluctance of females. The last constitutes, from the signaler’s perspective, neural noise in the receiver, analogous to amplifier noise in an electronic receiver. Such neural noise might result from limitations on attention
or memory or from a high threshold for response. For a male frog, communication is just as "noisy" regardless of the reasons for a female's failure to respond.

For a receiver, noise results from anything that contributes to error—in other words, anything that reduces the probability of an appropriate response to a signal. For a female frog, noise could result from heterospecific or suboptimal conspecific males with calls similar to those of optimal mates or from environmental attenuation and degradation of males' calls.

Note that Shannon and Weaver's use of the term "noise" corresponds to the signaler's perspective, whereas ordinary usage in general corresponds to the receiver's. In using this term, extra care is clearly needed. A common element in these two perspectives is the problem of environmental attenuation and degradation.

ENVIRONMENTAL ATTENUATION AND DEGRADATION

All signals eventually become inseparable from background energy in the environment at some distance from the source. This distance, the effective range of a signal, depends on the signal's intensity at the source, its attenuation and degradation during propagation through the atmosphere, and the level of irrelevant energy in the background. These influences are best understood for acoustic signals in terrestrial environments (Piercy and Embleton 1977; Wiley and Richards 1982; Gerhardt 1983; Michelsen and Larsen 1983). Effective ranges, or active spaces, have been determined for a variety of such signals: songs of red-winged blackbirds (Agelaius phoeniceus) (Brenowitz 1982); calls of blue monkeys (Cercopithecus mitis) (Brown and Waser 1984); and calls of a variety of frogs (Loftus-Hills and Littlejohn 1971) and insects (Römer and Bailey 1986).

An obvious way to increase the effective range of acoustic signals is to use frequencies not masked by background sounds. For instance, birds in a tropical forest in Panama appear to use frequencies lower than those of diurnal insects, mostly cicadas, in the same habitat (Ryan and Brenowitz 1985). Another way to extend the range of signals is to use bands of frequencies less susceptible to attenuation. For instance, in European marshes, Acrocephalus warblers emphasize frequencies in their songs that propagate best from the typical locations of their perches (Jilka and Leisler 1974).

The physical explanations for differences in the attenuation of sound are usually complex. Attenuation results from spherical spreading and, in addition, from atmospheric absorption, scattering, and interactions with the ground (Wiley and Richards 1982). The last is probably the least familiar of these effects to biologists. Propagation of sound between a source and a receiver that are both within centimeters of a porous surface with low acoustic impedance, like soil covered with a thin layer of vegetation, results in pronounced attenuation of frequencies above about 500 Hz (Embleton et al. 1976). This effect presumably explains why so many small insects and vertebrates avoid calling or singing from the ground when potential receivers are also on the ground (Paul and Walker 1979; Wiley and Richards 1982; Michelsen and Larsen 1983).

Transmission farther above the ground can result instead in pronounced attenuation of low frequencies, as a result of interference between the direct wave reaching the listener and the phase-shifted wave reflected from the ground (Piercy and Embleton 1977; Wiley and Richards 1982). This attenuation shifts to lower frequencies
as elevation above the ground increases. Most small birds do not produce frequencies low enough to be affected by this form of attenuation when singing more than about 1 meter above the ground. Primates in the canopies of tropical forests use frequencies much lower than those of birds in their long-range calls but still always above the band subject to attenuation by reflection from the ground (Waser and Brown 1984). The destructive interference between reflected and direct waves in these habitats results from the large phase shift as sound reflects from the porous soil, a surface of low acoustic impedance. Over a nonporous surface with high acoustic impedance, like water, low frequencies are not attenuated by interference in this way. As a consequence, birds in marshes use lower frequencies than those in grasslands (Cosens and Falls 1984; Wiley 1991).

In addition to reflection from the ground, atmospheric absorption and scattering from foliage attenuate sound in natural environments. Both of these forms of attenuation increase with frequency, and attenuation as a result of scattering from foliage increases more steeply with frequency in forests than in open areas (Morton 1975; Marten and Marler 1977). As a consequence, the frequencies used by birds for territorial songs in forests might differ from those used in open habitats. The expected differences, however, are not immediately clear. It is important to note that, above the band of frequencies affected by reflection from the ground, low frequencies attenuate least in all habitats. Thus, for songs with maximal range (or minimal effort for a given range), birds should always use dominant frequencies as low as possible for their body sizes. In forests, the attenuation of higher frequencies might result in a lower upper limit of acceptable frequencies. In fact, when body mass is controlled, birds in forests in eastern North America have lower maximal frequencies on average than those in open habitats. In contrast, with body mass controlled, there are no significant differences among major habitats in dominant frequencies (fig. 7.2; Wiley 1991). The narrower band of acceptable frequencies in forests thus influences the maximal, rather than the dominant, frequencies in songs.

Attenuation is not the only process that makes signals less distinctive to a receiver. Degradation of the temporal structure of signals can mask features that allow detection or recognition. Reverberation is a particularly important source of temporal degradation of acoustic signals. Among birds that communicate over distances of tens to hundreds of meters, the reverberations from trees in a forest obscure temporal structure within any frequency band. As a consequence, forest-inhabiting birds of North America tend to avoid rapid repetitions of frequencies in their long-range songs (fig. 7.3; Wiley 1991a). At long ranges in open habitats, sounds acquire irregular amplitude fluctuations as a result of refraction from moving cells of air that differ in velocity or temperature from the surrounding air (Richards and Wiley 1980). In these circumstances, birds produce songs of great temporal complexity, with rapid trills and other rapid repetitions of frequencies, often resulting in a tinkling quality (Wiley 1991a).

Similar considerations apply to communication with other sensory modalities. Vision in water, for instance, is limited by the pronounced frequency-dependent attenuation of light. Light attenuates within meters in water, rather than kilometers as in air. In either medium, objects that are distant in relation to the rate of attenuation lose contrast and take on the color and brightness of the general background. Underwater, a bright object near the surface is usually illuminated by downwelling light but is often viewed against the dimmer, horizontally scattered light called spacelight. The white tips on the wide fins of the epipelagic shark *Carcharhinus*
Fig. 7.2  Territorial passerine birds in eastern North America have songs that match the transmission properties of their habitats. (A) Within each of four broad categories of habitats (F, broad-leaved or mixed forest; P, parkland with scattered trees; S, shrubland; G, grassland), the dominant frequencies in songs of different species vary primarily with body size. (B) Maximal frequencies in songs, however, also differ significantly among habitats, presumably because lower attenuation of high frequencies in open grassland increases their effective range. Each graph includes the regressions of frequency on the inverse cube of body mass for species in the four categories of habitat; values for each species in forest (open squares) and grassland (solid circles) illustrate the variation within habitats (N = 29, 35, 21, and 18 for habitats F, P, S, and G, respectively). (See Wiley 1991a for statistical analysis.)

*longimanus*, for instance, resemble a school of fish when the shark's body is lost in the background spacialight. Presumably the shark thus lures its prey within striking range (Myrberg 1991).

Because short wavelengths both scatter best and attenuate least, optical signals at even modest distances or depths underwater become nearly monochromatic. At close range near the surface, yellow provides contrast with the dim blue spacialight and attenuates less than red. As a consequence, detection is improved by visual signals and receptor pigments offset toward longer wavelengths than the transmission maximum for water. Indeed, many fish in shallow water have receptor pigments with absorption maxima offset in this way (Lythgoe 1979, 1984). At a certain
Fig. 7.3 Songs of birds differ among habitats in inclusion of rapid amplitude modulation (indicated by sidebands) or short repetition intervals at any one frequency. In forested habitats, reverberation from foliage masks rapid recurrence of elements at the same frequency. F, broad-leaved or mixed forest; P, parkland with scattered trees; S, shrubland; G, grassland. (From Wiley 1991a, © 1991 by The University of Chicago.)

distance or depth, however, yellow becomes gray and red black. Dark/light contrast is the only effective way to encode signals in these conditions. The striking red and white patterns of some fishes that live at moderate depths or have crepuscular habits appear effectively black and gray. It may be easier for such species to acquire red pigments from the diet than to synthesize melanin, so that red pigments are the most efficient for making "black" patterns.

This discussion of environmental attenuation and degradation has focused on its consequences for simple detection of signals. These consequences also affect discrimination between similar signals. Signals that are distinct at their sources might become barely discriminable after attenuation and degradation.

COMPROMISES FOR SIGNALERS AND RECEIVERS

How can we understand the evolution of signaling, on the one hand, and responding to signals, on the other? Each involves its own compromises, as in each case it is not normally possible to maximize all objectives simultaneously. A brief overview of these compromises provides some orientation before the following section attempts a thorough examination.

The evolution of signaling should tend to maximize the probability of correct responses from the intended receivers. This trend, however, is subject to two constraints: efficiency in the production of signals, and risks from unintended receivers. The efficiency of a signal is appropriately expressed as a ratio: the probability of response by the intended receiver divided by the signaler's effort or risk (Wiley 1983). An efficient signal thus reliably produces a response in the intended receiver with a minimum of effort by or risk to the signaler. Unintended receivers are those that might intercept a signal and respond in a way disadvantageous to the signaler. For instance, a predator or rival might intercept a signal intended for a mate or offspring, to the signaler's disadvantage. A signaler must balance the advantages of responses to a signal by intended receivers against the effort and risks of producing a signal and the consequences of its interception by unintended receivers. Signals
should thus evolve to maximize the efficiency of obtaining responses by intended receivers and to minimize the probability of interception.

Receivers also face balancing advantages and disadvantages. Receivers must discriminate those signals to which a response is advantageous from those to which a response is disadvantageous. For each of the four basic associations of signals and responses (correct detections, missed detections, false alarms, and correct rejections), there is a net advantage (positive or negative) for the receiver. The mechanisms for discriminating signals should thus evolve to maximize the sum of the net advantages of each possible outcome times its probability. This sum is called the expected utility of the receiver’s responses.

**RECEIVERS’ ADAPTATIONS**

For a receiver, the object is to distinguish signals from irrelevant patterns of energy (or to distinguish different categories of signals) and then to associate each signal with the appropriate response. Because signals attenuate, degrade, and mix with irrelevant energy during propagation, some patterns of stimulation produced by a particular signal are not distinguishable by the receiver from irrelevant energy or from patterns produced by other signals.

The possibility of confusion is easily illustrated in terms of a simple signal discriminated from irrelevant masking energy by a single parameter, such as intensity. The masking energy often fluctuates randomly in intensity; the addition of a signal of constant intensity displaces this distribution toward higher, but often not distinctly higher, intensities (fig. 7.4). A receiver might respond or not depending on whether or not the intensity exceeds some threshold. This threshold then represents a simple criterion for response. In this case, the receiver can alter the probabilities of correct detection and false alarm (and consequently the probabilities of missed detection and correct rejection) by adjusting its criterion. Yet, as emphasized above, it cannot simultaneously maximize its probability of correct detection ($P_{CD}$) and minimize its probability of false alarm ($P_{FA}$). The possible compromises are represented

![Fig. 7.4 Derivation of the receiver operating characteristic (ROC) from the probability density functions (PDFs) for two signals (or signal plus irrelevant energy versus irrelevant energy alone) along a single stimulus dimension. For any criterion selected by the receiver ($T$), the probability of a correct detection ($P_{CD}$) equals the integral of one PDF from the threshold to infinity (shading with negative slope), and the probability of a false alarm ($P_{FA}$) equals the integral of the other PDF from the threshold to infinity (shading with positive slope). This pair of probabilities produces one point on the ROC, which is generated in its entirety by moving the threshold continuously from positive to negative infinity. The resulting ROC is a convex line lying above the positive diagonal in the unit square.](image-url)
Fig. 7.5 Receiver operating characteristics vary in shape as PDFs vary. Less overlap of PDFs (greater differences in means in relation to variances) results in a more convex ROC (compare plots on the left with corresponding plots on the right), one that reaches closer to ideal performance ($P_{CD} = 1.0$, $P_{FA} = 0$, the upper left corner). Deviations from normality in the PDFs result in variations in the shapes of the ROCs without altering their general properties.

by the locus of $P_{CD}$ as a function of $P_{FA}$, as a receiver's criterion varies continuously; this function is called the receiver operating characteristic (ROC, fig. 7.4).

The ROC for discriminations of this sort always lies on or above the positive diagonal and thus forms a convex set (Egan 1975; Macmillan and Creelman 1991). If the distributions of masking energy and of signal plus masking energy have the same shapes, then the ROC is symmetrical about the negative diagonal (fig. 7.5). Different variances or higher moments of the distributions produce an asymmetrical ROC. Furthermore, for all $P_{FA} > 0$, $P_{CD}$ increases monotonically with $P_{FA}$. Note that the greater the inherent detectability of the signal (the larger the difference in the means of the two distributions in relation to their variances), the more convex is the ROC. Conversely, an almost undetectable signal results in a nearly flat ROC, close to the positive diagonal.

This simple scenario easily generalizes to more complex situations. In particular, it applies to recognition of two signals with values that overlap along some dimension. Even when a signal and irrelevant masking energy, or two different signals, are distinctly separable at close range, all signals become less recognizable as they attenuate or as their temporal properties degrade.

This discussion has so far assumed that a response occurs whenever the receiver registers a value exceeding its criterion. This situation suggests a threshold in responses—in other words, an open-ended, all-or-nothing pattern of response as some parameter of a stimulus varies. In contrast, most open-ended patterns of response fit a more continuously increasing function of the features of a stimulus. Such patterns, often discussed in terms of a supernormal stimulus, occur widely among animals (Staddon 1975; Cohen 1984; Rowland 1989; Ryan 1990; Ryan, chap. 8). Increasing response as a continuous function of some parameter of a stimulus, however, does not necessarily indicate an underlying continuity in responsiveness.
Continuity could also result from a threshold for detection of a stimulus in noisy conditions. In this case signals with low values of the parameter would exceed the receiver's threshold less often than those with high values. As a consequence, despite an underlying threshold, any response to the rate of stimulation would change continuously as the stimulus varied.

An example of a threshold for response is provided by the treefrog *Hyla chrysoscelis*, whose mating calls differ primarily in pulse rate from those of its cryptic sibling species *H. versicolor*. In choices between two calls, female *chrysoscelis* prefer those with pulse rates above a temperature-dependent threshold, which corresponds to the pulse rate in calls of conspecific males at the same temperature (Gerhardt 1982). This threshold tends to inhibit responses to the wrong species, as *chrysoscelis* males at any temperature have higher pulse rates than do *versicolor* males.

Some patterns of response suggest tuned rather than open-ended preferences. In these cases, an optimal stimulus or range of values for a stimulus evokes a response. Female *H. versicolor*, unlike females of their sibling species, have preferences tuned to pulse rates similar to those of conspecific males at the same temperature (Gerhardt and Doherty 1988). The female green treefrogs mentioned earlier provide another example of tuned preferences (Gerhardt 1987).

Adjustments in tuning have much the same consequences as do adjustments in thresholds. If two different signals (or signals and masking noise) overlap along some dimension, then decisions about the tuning of responsiveness, like those about simple thresholds, affect the probabilities of correct detection and false alarm. Any tuning curve, behavioral or neuronal, is specified by its best value, its bandwidth, and its symmetry. When the properties of received signals vary, there is inevitably a trade-off between narrow tuning to reduce false alarms and broad tuning to reduce missed detections, analogous to the selectivity-sensitivity trade-off for electronic filters. Furthermore, continuous variation in any parameter of a tuning curve generates a receiver operating characteristic with properties like those discussed above (fig. 7.6). In particular, $P_{CD}$ is a convex function of $P_{FA}$, monotonically increasing for $P_{FA} > 0$. In addition, a more convex ROC results when two signals are inherently more discriminable (differ more in their means in relation to their variances).

A receiver adopts a criterion for responding to a signal by adjusting the parameters of a threshold or a tuning curve. The mechanism for this adjustment might be evolutionary, developmental, or physiological. This decision then affects the probabilities of the four possible outcomes for a receiver (correct and missed detections, false alarms, and correct rejections). How should this choice be made? Basic decision theory suggests that a criterion should maximize the expected utility for the receiver.

![Fig. 7.6](image)

Fig. 7.6 The principles that apply to thresholds for response also apply to filters or tuned responses. As the width of the filter narrows from $W-W'$ to $N-N'$, the probabilities of correct detection ($P_{CD}$) and false alarm ($P_{FA}$) fall along a convex line, just as in the case of variation in a simple threshold.
(see Sperling 1983 for a clear exposition of expected utilities and their applications to receiver operating characteristics). As a result of its imperfect performance in responding to a particular signal (or in responding differently to two signals), a receiver’s expected utility, $E(U)$, depends on (1) the rates of occurrence of different signals, (2) the vector of payoffs for the four possible outcomes, and (3) the criterion for a response, which sets the probabilities for the four outcomes.

Consider two signals (or a signal with background energy) that have different appropriate responses and occur with probabilities $\alpha$ and $1 - \alpha$, respectively. The probability density distributions of the two signals along some dimension separating them, together with the receiver’s choice of a criterion for response, determine the probabilities of correct detection ($P_{CD}$), missed detection ($1 - P_{CD}$), false alarm ($P_{FA}$) and correct rejection ($1 - P_{FA}$). Each of these four outcomes has an associated payoff: $h, m, a,$ and $j$, respectively. Each is a positive or negative net advantage. Thus the expected utility equals the sum, over all four outcomes, of the probability of each outcome times its payoff:

$$E(U) = \alpha h P_{CD} + \alpha m (1 - P_{CD}) + (1 - \alpha) a P_{FA} + (1 - \alpha) j (1 - P_{FA})$$

$$= \alpha (h - m) P_{CD} + (1 - \alpha) (a - j) P_{FA} + \alpha m + (1 - \alpha) j$$

The locus of values that yields constant expected utility is often called an indifference curve. To obtain such indifference curves in this case, we let $E(U)$ equal a constant, $U$, and rearrange the preceding equation, as follows:

$$P_{CD} = \frac{(1 - \alpha)(j - a)}{\alpha (h - m)} P_{FA} + \alpha (j - m) - j + U.$$

Each of these lines, like the ROC, represents the probability of correct detection ($P_{CD}$) expressed as a function of the probability of false alarm ($P_{FA}$). The slope is positive provided the payoff for a correct detection is greater than that for a missed detection ($h > m$) and the payoff for a correct rejection is greater than that for a false alarm ($j > a$). Both of these conditions are presumably met whenever it is advantageous for the receiver to respond to the signal.

Note that the slope remains constant but the intercept increases with increasing expected utility ($U$, above). We can therefore plot a family of parallel indifference lines on the same axes as the ROC (fig. 7.7A). The maximum realizable expected utility corresponds to the intersection of the ROC and the indifference line tangent to it. Therefore, for any ROC (determined by the probability density distributions of the features of signals and their alternatives) and for any expected utilities (determined by the probabilities of signals and alternatives and by the payoff vector for the four possible outcomes), maximum expected utility is attained by choosing a criterion for response that achieves $P_{FA}^*$ and $P_{CD}^*$ (the point of tangency in fig. 7.7).

The slope ($S$) of the indifference lines,

$$S = \frac{(1 - \alpha)(j - a)}{\alpha (h - m)}.$$

determines the point of tangency on the ROC and thus both the optimal criterion for response and the optimal combination of $P_{FA}^*$ and $P_{CD}^*$ (fig. 7.7B). For large $S$ (steep indifference lines), the optimal $P_{FA}^*$ and $P_{CD}^*$ are relatively small; for small $S$, the optimal $P_{FA}^*$ and $P_{CD}^*$ are relatively large.
Fig. 7.7 (A) Indifference lines plotted on an ROC reveal the combination of $P_{CD}$ and $P_{FA}$ that maximizes the expected utility and thus specifies (see fig. 7.4) the optimal criterion for response. (B) For any given ROC, indifference lines with high slopes ($S$) result in lower optimal combinations of $P_{CD}$ and $P_{FA}$. As a consequence, receivers with high $S$ should evolve adaptive gullibility or adaptive fastidiousness, criteria that result in many missed detections.

Consider two cases that present different problems of signal detection: detecting a subtle cheater (case 1); and searching for subtly discriminable objects (case 2), such as cryptic prey or an optimal mate. These cases serve to illustrate an important point: a receiver maximizing its expected utility should often not detect all deceivers nor respond to all optimal stimuli.

Case 1 applies to males that try to detect females but are parasitized by males that resemble females and occasionally steal matings. In this case, let the payoff for failing to detect a cheater equal 0 ($m = 0$), because the subject has now lost its advantage in obtaining a mate. Then the payoff for detecting a cheater equals the probability of finding a mate at some time in the future ($h > 0$); the payoff for correctly recognizing a female (rejecting the possibility of a cheater) equals the probability of successful mating once a female has been found ($j > h$); and that for a false alarm, in which the subject attacks a female mistaken for a cheating male, equals the probability of finding another female some time in the future ($a = h$ or perhaps 0). Under these conditions,

$$S = \frac{(1 - \alpha)(j - h)}{\alpha h}.$$

If the probability of finding a mate after detecting a cheater ($h$) is small compared with the probability of mating with a correctly identified female ($j$), then the slope ($S$) of the indifference lines is steep. It would pay for males to adopt a criterion for response to cheaters with low, but nonzero, $P_{CD}$ and consequently low $P_{FA}$. In other words, males in this case should not evolve to detect cheaters with complete accuracy. Instead, they should evolve some susceptibility to deception, a condition appropriately called adaptive gullibility.

Two other conditions also have this consequence: small $\alpha$, and a flat ROC. Evolution should favor susceptibility to deception (low $P_{CD}$) when cheaters are infrequent in comparison with actual females (large $[1 - \alpha]/\alpha$) or when cheaters and actual females are virtually indistinguishable (flat ROC, fig. 7.8). In the latter case, even a moderately steep slope for the indifference lines can favor complete gullibility, with $P_{CD} = P_{FA} = 0$.

Case 2 applies to most forms of searching behavior; for instance, a female attempting to choose an optimal mate from among many similar ones or a predator seeking cryptic prey. In these cases, a false alarm incurs the costs of selecting a suboptimal mate or prey, and a correct detection, the benefits of selecting an optimal
Fig. 7.8 For any given slope of indifference lines, a more convex ROC increases the probability of correct detections in relation to false alarms. (A) A signal that has low overlap with irrelevant stimulation and thus a highly convex ROC has high inherent detectability. (B) Steep indifference lines in combination with a nearly flat ROC can favor a criterion that excludes any response. Such signals are best ignored.

one ($h < a$). A missed detection (or a suitable mate or prey) or a correct rejection (of a suboptimal object) entails additional search, with its associated costs of time lost and risks encountered. We can think of these costs as devaluing the expected utility at the next opportunity for a response ($U$) by a factor $\lambda$. Under these conditions,

$$S = \frac{(1 - \alpha)(\lambda U - a)}{\alpha (h - \lambda U)},$$

where $h > \lambda U$ and $1 > \lambda > 0$. Here the steepness of the slope, $S$, increases with $\lambda$, which varies inversely with the costs of additional search. So low search costs favor a criterion that yields low $P_{CD}^*$ and low $P_{FA}^*$. In other words, searches should evolve to pass many suitable objects of their choice (low $P_{CD}^*$), a tendency appropriately called adaptive fastidiousness. Real (1990) reached a similar conclusion in discussing the theory of sequential search.

Again two other conditions favor low $P_{CD}^*$ and $P_{FA}^*$: small $\alpha$, and a flat ROC. For any constant cost of search, fastidiousness is favored when optimal objects of search are infrequent (large $[1 - a]/a$) or when optimal objects have low inherent detectability (flat ROC, fig. 7.8).

Cases 1 and 2 represent two views of situations that generate indifference lines with steep slopes. In both cases, maximizing expected utility leads to low $P_{CD}^*$ and low $P_{FA}^*$. Case 1 emphasizes the advantages of gullibility in response to suboptimal signals; case 2 emphasizes the advantages of fastidiousness in response to optimal signals. Gullibility and fastidiousness are two aspects of the same problem.

**SIGNALERS' ADAPTATIONS**

Signalers face different problems depending on whether or not it is in the receiver’s interest to respond. In cases of manipulation or deception, signalers receive an advantage from a response to a signal that is disadvantageous to the receiver. In these cases, the signaler must maximize the probabilities of a false alarm by a receiver. This objective is attained by producing a signal that mimics another signal (the model) that evokes the desired response in appropriate circumstances for the receiver. Mimicking a female to take advantage of a more successful male’s attraction
of females, simulating the presence of several rivals rather than just one (a Beau Geste effect: Krebs 1977, but see Yasukawa 1981), and mimicking an alarm call to trigger withdrawal of an opponent all fall into this category. Maximizing a response from the receiver in these cases, as demonstrated in the discussion of case 1 above, depends in general on close mimicry of the model, to achieve low inherent detectability by the receiver and thus a flat ROC, and on infrequent use, to assure low $\alpha$.

When it is advantageous for the intended receiver to respond to a signal, then signalers should evolve to maximize the probability of correct detections by the receiver and to minimize disadvantageous interception by unintended receivers. The best way to increase a receiver's correct detections depends on its search costs, as described above for case 2. When the receiver's search costs ($\lambda$) are low enough that the indifference lines have slope $S - 1$, then frequent signaling (high $\alpha$) and high inherent detectability (convex ROC) favor increased correct detections by the receiver. On the other hand, when the receiver's search costs are high enough that $S < 1$, then increased detectability could instead favor lower probabilities of correct detection.

The inherent detectability of signals becomes a problem especially in situations with many similar signals or with high attenuation and degradation of signals. In such situations, there are four basic ways to increase the inherent detectability of signals: (1) increasing the intensity of signals, (2) increasing their contrast with other patterns in the environment, (3) increasing the spatial or temporal stereotypy (redundancy) of signals, and (4) increasing the possibilities for the receiver to act as a matched filter (Wiley 1983, Harper 1991). For example, calling male frogs, to increase the chances of a female's response, could (1) evolve increased intensity of calling, (2) shift to frequencies or times with less masking sound, (3) repeat the call more often, and (4) use a stereotyped frequency (or frequency-temperature relationship) in stereotyped temporal patterns and at stereotyped locations. In general, frogs exemplify these adaptations.

The first two of these adaptations increase the signal's contrast with irrelevant energy. Increases in intensity, whether louder sounds, larger or brighter visual stimuli, or greater quantities of a pheromone, beyond some point require changes in morphology or synthetic pathways. These changes in turn require compromises with other demands on the signaler. Shifts to new parameters not masked by irrelevant energy in the environment require a correlated shift on the part of the receiver. In the initial stages, the advantage of the new signal to a receiver as a result of increased detectability would have to balance any disadvantage of additional search time for a rare signal. On the other hand, if a receiver at first had a more broadly tuned receptor, rather than one with the best stimulus shifted, then this variant receiver and the variant signaler might both realize immediate advantages.

Increased redundancy in signals has a well-known influence on the probability of correct detection (Shannon and Weaver 1949; Wiley 1983). Redundancy can include both spatial and temporal relationships among the components of a signal. In either case, predictable relationships among the components of a signal allow a receiver to reconstruct the correct signal from an imperfectly received one. The receiver thus has more than one chance for correct detection of the signal. Redundancy in animals' displays for communication is thus closely related to the classic ethological concept of ritualization (Cullen 1966; Wiley 1983).

Stereotyped temporal and spatial relationships among the components of a signal have the disadvantage of reducing the amount of information encoded in signals in unit time (or area). Consequently, these signals find their application in
situations in which the probability of error in responses is high but the advantage of diverse or variable signals is low (Wiley 1983).

Receivers can detect signals most reliably when they are maximally prepared for them. For instance, the ability of human observers to detect signals in masking noise is improved by giving them only one, rather than several possible, signals to detect (reviewed by Wiley 1983). When a signal must be learned, those with easily remembered parameters enhance detection (Guilford and Dawkins 1991). In addition, human observers perform better when alerted about the time interval and the location in which a signal might occur. For instance, a light before or just after the possible occurrence of a sound in masking noise makes detection of the sound more reliable. In this case, the light serves as an easily detected alerting signal that specifies the time at which a less easily detected signal might occur. Such an alerting signal would also improve recognition, which involves classification of signals in addition to simple detection (Richards 1981a; Wiley and Richards 1982; Wiley 1983).

These improvements in the receiver’s performance as a result of its knowledge of the possible occurrence or exact parameters of a signal could have two explanations. For electronic systems, such prior knowledge could be used to develop a filter optimally matched to the expected signal. Animals might conceivably alter the filtering characteristics of their sensors as well. Sensitization to repeated stimuli suggests such an effect. The improvement might also result from greater attention focused on the task at hand. Indeed, attention, in psychology, seems analogous to matched filtering, in engineering. Provided it is advantageous for receivers to detect or to recognize a signal, signaling should evolve to provide opportunities for attention or matched filtering, and receivers in turn should evolve to take advantage of these opportunities.

Unintended receivers can constrain the evolution of efficient signals. In some cases, special properties of signals can limit their detection to intended receivers. Use of high-frequency sounds, for instance, limits potential receivers to those in the immediate vicinity of the signaler and thus reduces possibilities for detection by predators or rivals. This consideration might explain their use in some alarm calls, juvenile begging calls, and precopulatory calls of birds. Low intensity also results in a relatively small active space, but high frequencies have the advantage of allowing high intensity close to the source in combination with a small active space. A particularly striking case of “private” communication occurs in a deep-sea fish, in which the photophores produce red light that has a high attenuation rate in water but is exactly matched by the species’ receptor pigments (O’Day and Fernandez 1974).

**IMPLICATIONS OF ERRORS IN COMMUNICATION**

To explore possible adaptations for reducing errors in communication, the remaining sections consider the evolution of (1) behavioral boundaries between species, (2) spectacular displays, and (3) honesty and deception. These topics shed some light on the evolution of communication in the cases described in the opening section: mate choice by female frogs; interactions of territorial birds by singing; and deceptive use of alarm calls. The discussion emphasizes how little we know about errors in communication despite their central place in explaining the evolution of receivers’ and signalers’ behavior. A recent controversy concerning the nature of species provides a start.
Evolution of Behavioral Boundaries between Species

Do populations evolve adaptations for recognizing suitable mates or for rejecting unsuitable ones? Recent discussions of species limits have disagreed over whether species are more consistently characterized by recognition of conspecifics or by isolation from heterospecifics (Patterson and Macnamara 1984; Endler 1989). In other words, does the evolution of mate choice result in acceptance of conspecifics or rejection of heterospecifics? This distinction is obscured when a receiver's task is viewed as a problem in signal detection. As we have seen, probabilities of correct detection (response to a suitable mate) and false alarm (response to an unsuitable mate) do not usually admit independent adjustment. Consequently, probabilities of correct detection and correct rejection (rejecting an unsuitable mate) cannot usually be maximized simultaneously. To permit such simultaneous maximization, conspecific and heterospecific signals would have to be completely distinct to the receiver's sensors despite attenuation and degradation.

If hybridization is disadvantageous, selection should favor reduction of false alarms in mate choice. This situation thus fits case 2 above, with \( a < 0 \). With steep indifference lines, high probabilities of correct detection combined with low probabilities of false alarms can only occur with a highly convex receiver operating characteristic, a result of signals with high inherent distinctiveness. Consequently, signals might evolve greater distinctiveness in order to reduce either the frequency of dysgenic hybridization or the time required to find a conspecific mate. Scenarios for the divergence of mating signals thus fall between two extremes: either (1) the two populations initially hybridize after contact but subsequently diverge in their signals for mate choice as a result of selection against hybridization; or (2) the populations do not hybridize at the outset as a result of differences in the final (often short-range) signals for mate choice but subsequently diverge in their initial (long-range) signals as a result of selection for reduced time spent searching for a mate. In each case, selection for divergence would have to come from advantages of increased inherent discriminability of conspecific signals. Of course, populations in the first situation might evolve into the second situation if final (short-range) signals diverged before initial (long-range) signals did. Note that once initial signals are distinct, there is little or no selection for divergence of final signals for mating.

Few, if any, studies of hybrid zones have provided evidence specifically for the first scenario, often called reinforcement of reproductive isolation in sympathy (Butlin 1989). Several zones of contact between species of frogs provide evidence for greater divergence of signals in sympathy than in allopatry (Littlejohn 1965, 1981; Fouquette 1975; Ralin 1977), although whether each of these cases represents the outcome of the first scenario above (reinforcement) or the second (often called reproductive character displacement) remains uncertain.

Divergence of signals might not occur if changes in the properties of signals brought disadvantages in propagation. If so, these disadvantages in attracting conspecifics would cancel any advantages of attracting proportionately fewer heterospecifics. Reinforcement might also occur by changes in females' preferences without any change in signal properties. To increase selectivity for conspecific signals (decrease false alarms in proportion to correct detections), receivers could raise their thresholds (or narrow their tuning) for responses. These changes would also be likely to increase search time (a result of increased missed detections), so that advantages of selectivity would balance disadvantages of additional search. These considerations make it clear that any investigation of the evolution of behavioral boundaries between species must consider the net advantages of changes in signals and receiv-
ers' criteria in relation to the probabilities and consequences of false alarm, correct rejection, missed detection, and correct detection.

Mimicry of mating signals by potential predators drastically alters the optimal probabilities of correct detection and false alarm for the receiver. In the best-known case of this possibility, male *Photinus* fireflies responding to conspecific females' flash patterns must avoid the similar flash patterns of predatory *Photuris* females of several species (Lloyd 1981, 1985). In this case, the low inherent discriminability of conspecific from predatory signals results in a relatively flat receiver operating characteristic, and the high cost of false alarms in relation to the benefits of correct detections, as in case 2 above, produces steep indifference lines. Thus this situation should favor receivers with adaptive fastidiousness: low probabilities of correct detection as a result of high thresholds (or sharp tuning) for response. The males' cautious approaches to flashing females suggest such a situation. Female *Photuris* lure males of several species of *Photinus*; apparently the difficulty of mimicking several species provides opportunities for cautious *Photinus* males to detect most such traps. Female *Photuris* in turn appear to leave themselves open to conspecific males that mimic the flashes of their prey and thus locate them before they are ready to mate. It would be fascinating to have a full analysis of signal detection by these fireflies, including estimation of the expected utilities of response criteria.

### Evolution of Spectacular Signals

Elaborations of displays that result in increased intensity, contrast with irrelevant patterns of energy, and temporal and spatial redundancy all serve to improve an intended receiver's performance by increasing the inherent detectability of the signal. Such signals thus offer advantages to signalers when attenuation and degradation are high or when many different signals require discrimination, so that receivers would otherwise miss many signals. We can call signals that have evolved in this way spectacular displays; their defining characteristic is high inherent detectability (or discriminability from other signals).

Previous explanations for the evolution of elaborate displays fall into two general categories: (1) runaway sexual selection for exaggerated features of signals as a result of arbitrary female preferences (those with no consequences for the female's survival or reproductive success or for her progeny's success) (Fisher 1930; Lande 1981; Kirkpatrick 1987); and (2) selection for costly signals to ensure honesty (Zahavi 1975; Kodric-Brown and Brown 1984; Andersson 1982a, 1982b; Grafen 1990a, 1990b; Maynard Smith 1991). Exaggerated or costly displays are, however, not necessarily spectacular, in the sense of being inherently highly detectable. Basic considerations of signal detection suggest, however, that inherent detectability is a crucial consideration.

For females, the consequences of choosing the wrong mate can include less than optimal genes for her progeny or less assistance in raising her offspring. If the cost of a false alarm (choosing a suboptimal mate) is high (net advantage low) and the cost of a missed detection (additional search for an optimal mate) is low (net advantage high), this situation resembles case 2 above. Steep indifference lines and a shallow receiver operating characteristic, as a result of signals with low inherent discriminability, favor the evolution of adaptive fastidiousness. Because the low probabilities of response in this case create "noise" for a signaler, signaling by optimal males should evolve toward more spectacular displays.

It is important to emphasize again that fastidiousness of a receiver in itself creates "noisy" communication from the viewpoint of the signaler (see above). For
a signaler, the situation is exactly the same whether receivers fail to respond to signals because of environmental attenuation or because of "neural attenuation," a result of high thresholds or narrow filters for response. In either case, more spectacular signals—ones with greater intensity, redundancy, and distinctiveness—are advantageous (Wiley 1983).

From the female's perspective, the difficulty of detecting optimal mates presumably increases when she has little time to make her choice, when she has little experience with potential mates, or when she responds to attenuated signals from potential mates at long distances (Wiley 1983). These three situations, respectively, reduce the redundancy from prolonged communication with potential mates, make mates of different quality inherently less discriminable, or require discrimination of attenuated and degraded signals. Females should encounter greater problems in optimal mate choice when their previous experience with these signals is limited, for instance, when young are reared by females alone, when sexes do not associate regularly except during mating, and when mate choice occurs quickly. These situations favor a shift in females' preferences toward signals with greater inherent detectability (or discriminability). Thus when a response is made after a short time, with little experience, or at long range, a female's preference for spectacular signals has advantages in reducing false alarms and missed detections. Such a preference is adaptive rather than arbitrary.

The process of runaway sexual selection, as a result of genetic correlation between signal properties and female preferences, can occur in theory with completely arbitrary female preferences. Nevertheless, the conditions for sexual selection depend strongly on the net advantages of female preferences (Heisler 1984; Pomiankowski 1987, 1988). From the preceding analysis of female preferences as a problem in signal detection, it seems likely that these preferences are rarely, if ever, completely arbitrary. A full analysis of these advantages must incorporate both the probabilities of false alarms, missed detections, and correct detections and their consequences (suboptimal mating, additional search, and optimal mating, respectively) for a female's reproductive success.

Any shift in a female's preferences for mates seems likely to alter the expected utility of her criterion for response, as derived above. Direct selection for female preferences, such as selection for an increase in the expected utility of the criterion for detecting optimal mates, makes the evolution of adaptive preferences more likely, because it lowers the threshold required for runaway evolution (Heisler 1984). Thus the universality of the problems of signal detection tends to favor the evolution by sexual selection of exaggerated signals that are also inherently detectable or discriminable—in other words, spectacular displays.

The distinctive acceleration of sexual selection requires genetic correlation as a result of assortative mating between signalers and receivers. This acceleration, despite Zahavi's (1991) recent arguments, can thus only apply to signals that affect recognition of mates. Otherwise, signals and receivers' criteria no doubt still co-evolve, in the sense that the advantages of any signal depend on the criteria of receivers in the population and, conversely, the advantages of any criterion depend on the available signals, but the acceleration as a result of sexual selection is absent.

Limitations on the evolution of spectacular displays include the disadvantages to signalers from interception of their signals by unintended receivers, often predators, parasites, or rivals. Bright coloration increases predation in fishes (MacPhail 1969; Endler 1980, 1983); calling attracts parasites to crickets (Cade 1979) and predators to katydids and frogs (Bellwood and Morris 1987; Ryan et al. 1981). These
disadvantages of spectacular displays, however, are often partially mitigated by mechanisms for concealment when not in use. Many lekking birds with spectacular plumage can conceal conspicuous features when not displaying (Wiley 1991b). Túngara frogs (*Physalaemus pustulosus*) utter the wideband component of their calls, the "chuck," primarily when calling in groups; this component attracts both females and predaceous bats (Ryan et al. 1981; Rand and Ryan 1981; Ryan 1983). Thus, by adding "chucks" when calling in groups, males increase their mating success in competition with other males and minimize the consequent risks of predation.

Spectacular signals also have disadvantages in rapid transmission of information, as a consequence of their high spatial and temporal redundancy. Thus situations that require high rates of transmission of information, such as negotiations during close-range interactions of mates or opponents, should favor diverse and variable signals rather than stereotyped ones. This contrast is evident in several comparisons of close-range and long-range communication in birds and primates (Marler 1973; Wiley 1973). The evolution of spectacular displays is thus constrained not only by risks of predation or parasitism but also by limitations on rates of transmission of information.

**Evolution of Honesty and Deception**

Sexual selection, when female preferences are adaptive, is a special case of the evolution of honesty in signaling. Much as in the case of sexual selection, the evolution of honesty in general must often result from the evolution of receivers' preferences for inherently reliable signals. Simple signals, easily bluffed or mimicked, tend to become unreliable. For honest signalers, it then pays to escalate the intensity or persistence of display until imitators can no longer match them (Andersson 1982a, 1982b; Kodric-Brown and Brown 1984). For receivers, it is advantageous to respond selectively to these escalated signals or to any others that preclude mimicry or bluffing. In the case of signals indicating overall strength or vigor, costly displays or those directly revealing vigor should prevent imitation by less vigorous signalers. Either way, the result is honest signals.

When communication is viewed in terms of signal detection, the conditions for the evolution of honest, costly signals nearly match those for the evolution of spectacular, exaggerated displays. Both sorts of signals can result from receivers' objectives of increasing correct detections and decreasing false alarms. These objectives can be achieved in any one, or a combination, of at least three ways: receivers can (1) adopt more stringent criteria, (2) switch to inherently more reliable signals, or (3) probe the honesty of signalers at irregular intervals (Wiley 1983). Each possibility raises additional issues.

If receivers evolve more stringent criteria for response, honest signalers should evolve progressively more exaggerated displays. Exaggeration, spectacular display, and high cost might thus often evolve together. When a spectacular, exaggerated signal requires overall vigor, its high cost results in high inherent detectability for a receiver, because deceiving signalers cannot match the costly signal. Note that it is not enough for such a signal to have high cost or exaggeration; it must also be spectacular, so that a receiver can easily discriminate it from less costly displays. Otherwise, there is no increase in the receiver's expected utility from increasing its criterion for response.

If a receiver changes its criterion to reduce false alarms (from bluffs or mimicry), it also usually increases its missed detections. Consider a newly arrived migrant bird looking for a territorial vacancy. If it adopts a lax criterion for judging whether
an area is already occupied by a serious defender, it risks false alarms, cases in which it fails to challenge minimal advertisement by a weak opponent or by an opponent trying to claim an oversized territory, and thus incurs additional costs of search or risks failing to find any territory. On the other hand, if it adopts a stringent criterion, it risks missed detections, cases in which opponents fight vigorously when challenged. Thus a more stringent criterion (higher threshold or narrower tuning) encounters both the advantage of fewer false alarms (vacancies passed up by mistake) and the disadvantage of more missed detections (challenges that are lost causes). The balance of advantages and disadvantages then determines whether or not it pays to adopt more stringent criteria for response to a signal and hence whether or not exaggeration of the signal occurs.

A second alternative for receivers, instead of a more stringent criterion for response to an unreliable signal, is a switch to a signal with less bluffable parameters, ones that inherently permit discrimination of honest and deceptive signalers. The intensity of the carotenoid pigments of male guppies (Poecilia reticulata), as an indicator of feeding efficiency provides a possible example of such a parameter (Endler 1980; Kodric-Brown 1989). In this case honest signals need not be exaggerated to assure high detectability.

A third alternative for receivers subject to deception is to probe signalers by calling their bluff. If bluffers or mimics, once discovered, incurred high losses, then these disadvantages of bluffing could promote honesty in signaling. On the other hand, as any gambler knows, calling a strong opponent’s bluff has its own costs. The consequences for exaggeration of signals seem complex. If probing occurred frequently, honesty would tend to spread at the expense of bluffing. On the other hand, probing would tend to spread only if bluffing occurred frequently. These reciprocal influences could lead to evolutionarily stable mixed strategies of probing and gullible receivers, on the one hand, and honest and bluffing signalers, on the other. In general, probing could stall the evolution of exaggerated signals in response to bluffing and mimicry by assuring high inherent detectability of simple signals.

This discussion emphasizes that, as a result of the interaction of selection on signalers and receivers, signals normally evolve toward honesty. The problem in the evolution of deception is thus the nature of limitations on the evolution of honesty. Since deception relies on errors by receivers, selection should always favor receivers that minimize deception by increasing discrimination, shifting to more reliable signals, or probing. However, receivers face trade-offs that can limit their adaptations to deception.

The theory of signal detection makes these trade-offs clear (Wiley 1983). As explained above, a receiver’s performance can be understood only in terms of the inherent detectability of signals (and consequently the receiver operating characteristic) and the criterion for response. The expected utility of any criterion in turn depends on the probabilities and net advantages of each of the four possible outcomes of a receiver’s decisions to respond or not. Dawkins and Guilford (1991) have also emphasized the importance of receivers’ costs in explaining the evolution of deception. These costs are incorporated in the expected utility of a receiver’s criterion for response. As shown above, maximizing this expected utility can lead to evolutionarily stable deception.

Such adaptive gullibility can, for example, explain the evolutionary stability of deception by males that mimic females in order to steal matings. For males intrinsically more likely to attract females, the costs of false alarms (attacking an actual female mistaken for a deceptive male) can make it advantageous to accept some
deception (missed detections of female mimics). Thus, depending on the payoff vector and probabilities of true and false signals, the receiver's expected utility is maximized by accepting some level of deception rather than by changing its criterion for response in order to minimize deception. In this case, deception becomes an evolutionarily stable feature of communication.

In the case of deceptive alarm calls, the situation is similar. Here the costs of missed detections (failing to respond to a true alarm) make it advantageous for a receiver to accept some deception (false alarms). Again, the receiver maximizes its expected utility by a criterion that results in some deception, and thus deception becomes an evolutionarily stable feature of communication.

EQUATIONARY STABLE DECEPTION

In a series of elegant expositions, Grafen (1990a, 1990b, 1991) presented a strong case for the evolution of honesty in signaling. In his model of communication, the evolutionarily stable strategy for receivers is to respond only to reliable signals. Consequently, the evolutionarily stable strategy for signalers is to produce such signals. He concludes trenchantly, "Receivers must get what they want in a stable signalling system" (Grafen 1990a, 526). The model of communication analyzed by Grafen involves advertisement and assessment. A signaler's objective is to persuade the receiver that its quality is as high as possible, and the receiver's objective is to evaluate the signaler's quality.

In this model, the deduction that honesty in signaling is evolutionarily stable requires only a few assumptions. Most important is the assumption of "continuity in everything" (Grafen 1990b, 476): the signaler's fitness is a continuous function of its actual signaling level, its perceived signaling level, and its quality. Reliable or honest signals require costs for signaling, and in particular, greater costs for signalers with lower quality. Provided there is continuity in everything, it does not pay to try deception: for any increase in signaling, above the signaler's evolutionarily stable level, the gain from enhanced perception by the receiver is more than compensated by the increased cost. Because every signaler has its own equilibrium related to its quality, honesty in signaling is the only evolutionarily stable strategy.

In his discussion of the limitations of this model, Grafen (1990a: sections 5 and 6) considered two possibilities that might restrict honesty in communication. One of these possibilities, the most obvious element of the present approach missing in Grafen's models, is error by receivers. His brief discussion of this issue (Grafen 1990a, 528) makes it clear, however, that imperfect perception by receivers, short of outright blindness, has little influence on his deductions. The assumption of "continuity in everything" assures that an error-prone receiver's perceptions are continuously related to expectations of the signaler's quality. Receivers' errors alone are not sufficient to limit honesty in signaling.

The situation changes, however, if we abandon the assumption of continuity. As Grafen notes, when he considers this second possible restriction on honesty, there might exist an alternative set of signalers or modes of signaling for which "signals are much cheaper for a given quality [of the signaler]" (Grafen 1990a, 533). This possibility of cheap imitations of honest signals, in combination with error by receivers, produces the conditions for evolutionarily stable-deception.

Such deceptive signals cannot occur too frequently in comparison with honest signals, otherwise receivers would do better to ignore the signals. A receiver unable
to discriminate between deceptive and honest signals must rely on the average gain when deciding to respond (Wiley 1983, Grafen 1990a, 534–35). The present treatment of communication has clarified this condition. To maximize its fitness, a receiver must adjust its criterion for response to maximize its expected utility, a quantity that depends on the frequencies of honest and deceptive signals and on the consequences for fitness of responding or failing to respond to each.

Grafen's analysis is thus in substantial agreement with the present approach. As a general rule, the interaction between signaler and receiver leads to honesty in communication. Evolutionarily stable deception is an exception to the rule. Grafen's analysis clarifies the importance for the evolution of stable deception of some discontinuity in the relations between signals or costs and signalers' states. The present analysis emphasizes the importance of imperfect discrimination by receivers. This condition in turn requires an evolutionarily stable limitation on discrimination. The application of decision theory to communication makes this limitation clear: maximizing the receiver's expected utility in responding to a signal can limit the advantages of discrimination by a receiver and, provided there is not "continuity in everything," can lead to evolutionarily stable error and deception.

CONCLUSION

Errors by receivers, in the sense of evolutionarily inappropriate responses to signals, are likely to occur during communication at long range, in dense aggregations, or in the presence of deception. Such errors have major implications for the evolution of communication. For signals that cannot be completely distinguished by a receiver, there are four possible results of any decision to respond or not to respond to a signal: correct detection, missed detection, false alarm, and correct rejection. In general, the probabilities of these results cannot be independently adjusted by a receiver; in particular, criteria for response that increase the probability of correct detection also inevitably increase the probability of false alarm. The inherent discriminability of the signal, together with the probability and net advantage of each outcome, determines the expected utility of any criterion for response. Maximizing the expected utility of a receiver's criterion can lead to adaptive gullibility (evolutionarily stable susceptibility to deception) or to adaptive fastidiousness (low responsiveness to signals). In the latter case, signalers must contend with increased uncertainty in the responses to a signal. Signalers in such situations can improve signaling efficiency by increasing the inherent discriminability of signals, often by exaggeration. Female choice of mates, as an example of adaptive fastidiousness, can result in such exaggeration of signals. In general, explanations for the evolution of any receivers' performance, as in the cases of female choice or evolutionarily stable susceptibility to deception, require evaluation of the expected utility of the receivers' criterion as well as the probability and inherent detectability of the signal.

ACKNOWLEDGMENTS

I thank members of the Animal Behavior Seminar and Ecolunch at Chapel Hill for thorough review of the ideas expressed here. Many helpful suggestions were also received from Carl Gerhardt, Michael Ryan, Lori Wollerman, Jean Boal, and Helmut Mueller.
REFERENCES


