Noise in the Evolution of Communication and Thought

How Natural Selection and Noise Shape Animal and Human Minds

Noise in the Evolution of Communication and Thought

How Natural Selection and Noise Shape Animal and Human Minds

R. Haven Wiley

Hickory Woods Press
Carrboro, NC

2023

Noise in the Evolution of Communication and Thought:
How Natural Selection and Noise Shape Animal and Human Minds

R. Haven Wiley
http://rhwiley.bio.unc.edu

Layout and cover design by the author.

Published by Hickory Woods Press, Carrboro, NC 2023
Printed in The United States of America

First printing

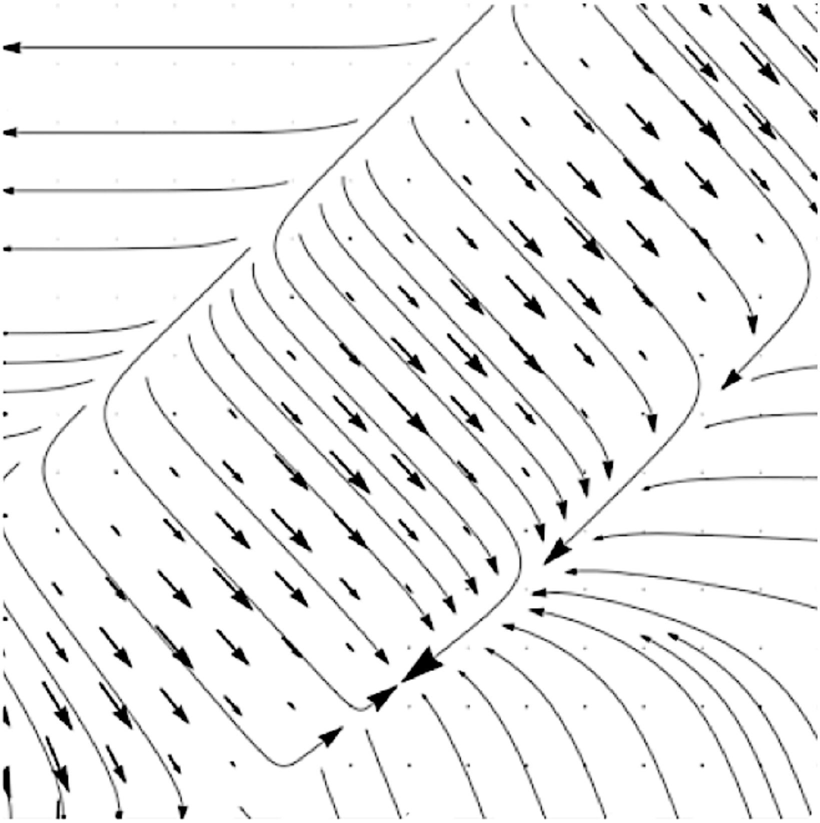
For additional copies visit
www.amazon.com/stores/R.-Haven-Wiley/author/B0116DR6R4

Library of Congress Control Number: 2023905624

ISBN 978-1-7325227-1-8

By the same author

Noise Matters: The Evolution of Communication
Harvard University Press, 2015



Frontispiece. Possible paths of evolution by natural selection for the relationship between the level of a signaler's exaggeration of a signal (horizontal axis) and the level of a receiver's threshold for a response to the signal (vertical axis).

See page 178

Contents

Section I. From Communication to Cognition

Section II. Difficult Issues in Communication

Figures and references are presented at the end of each chapter.

For the many students at the University of North Carolina-Chapel Hill
whose enthusiasm and curiosity helped to move these ideas along.

Preface

The chapters of this collection have all been published previously in separate places, but together they make a continuous story. This story is the evolution of organisms in the presence of noise. Evolution by natural selection produces the adaptations of organisms to their environments. These adaptations include both those to organisms' external environments and those to their social environments. Communication is one of the adaptations to social environments.

My book, *Noise Matters: The Evolution of Communication in Noise,* explains how the presence of noise makes a big difference in the evolution of adaptations for communication. It also reveals that the same principles apply to the evolution of perceptions of sensory inputs. Furthermore, it shows that evolution of communication and perceptions, when noise is present, affects how organisms think.

Each of the preceding statements might seem either self-evident or preposterous, depending on your preconceptions. Nevertheless, regardless of how they might appear to you, it turns out that the evolution of communication in noise leads to some surprises. *Noise Matters* attempts to introduce the path that leads to some of these surprises. The essays in the present collection start from different approaches, and they go farther.

My objective is to develop a view of communication in the presence of noise. This approach, it becomes apparent, leads to a reassessment of long-prevailing definitions of communication, signals, and receivers. Communication requires signalers, those who produce signals, and receivers, those who respond to them. Taking noise into account requires re-thinking what signals and noise are and how signalers and receivers relate to each other.

Just as important, these essays develop a new view of communication as evolutionary adaptations, something that requires living organisms subject to evolution. The mechanisms of evolution – mutation, drift (randomness in demography), migration, and natural selection – are all mathematical in essence. Natural selection is the one that produces adaptations to current circumstances. It is the action of natural selection on communication in the presence of noise that generates a new view of communication.

This exploration of communication eventually leads to some wide vistas. If our communication as writer and reader succeeds, it will reveal the general honesty of communication, warn of inescapable dishonesty, indicate ways to minimize miscommunication, illuminate complexities of human language, disclose the fundamental similarities between humans and other organisms, suggest why even surprisingly intelligent animals still lack language, show the similarity between communication and perception, lead to new understandings of free decision, self-awareness, and thinking, and finally locate the basis of human morality in an ancient adage.

In Section I essays introduce the evolution of communication in noise in different ways and then show how the results apply to the evolution of human language and cognition. They confirm that noise makes all organisms into decision-makers. Choice or decision is not restricted to those with advanced mentality, such as humans, but occurs (indeed, must occur) in even the simplest organisms, even in plants. Advanced communication, such as human languages, require more complexity but, nevertheless, might evolve in surprising ways. Other cognitive features, self-awareness and freedom of will or decision, both with long histories of philosophical debate, have surprisingly intuitive explanations. The consequences of evolution in the presence of noise reach into every corner of our ways of thinking.

The final essay in Section I scrutinizes natural selection, the fundamental process in the evolution of adaptations. Because it produces such surprising results for the way we humans think, it is best to examine exactly what natural selection involves. This examination reveals just how far natural selection can affect our thinking.

In Section II of this collection, the essays continue to examine issues fundamental for adaptations of communication in noise. The first of these essays presents the mathematics of the relevant natural selection. Then the second essay explores the procedures of Signal Detection Theory, the framework for understanding communication and perception in noise. The third addresses complexities in thinking about transfers of information by communication. Finally, a summary wraps up all of these threads.

Each of these essays, in both sections, leads to surprising results. Each requires some adjustments to preconceptions.

Section I

Introduction
From Communication to Cognition

This first section presents six essays that introduce, without technicalities, the problem of how communication evolves in noise, and then they develop consequences for the nature of human language, perception, self-awareness, and free will. Evolution by natural selection provides the fundamental mechanism for this understanding of communication and perception in noise and for all of the consequences. The final essay in this first section thus considers natural selection more thoroughly. It produces living organisms that must lead their lives in what is, for them, a noisy universe. It has produced organisms that have proposed philosophies of biology – and it proposes the biology of philosophy.

Even a start on our path to explore communication requires some care. After all, what is communication? To get started, we must agree that communication occurs when a receiver responds to a signal from a signaler by a change in behavior. To be more precise, a receiver's behavior after a signal must differ statistically from its behavior before. A receiver need not behave differently to every single instance of a signal, but it must change in some way more often than expected by chance. A receiver must, as a rule, be different after a signal than before.

The change in a receiver can be an immediate overt response, but it need not be. Any change, even if delayed or covert, counts. For instance, answering questions correctly following instruction counts. Psychological disturbance after traumatic experience counts. Memory counts. Indeed any change in a receiver's nervous system, or even more generally in its constitution of any sort, counts as communication. The only requirement is that this change must occur in response to a signal. In other words, it must be associated, in a statistical sense, with a signal.

Next, what is a signal? A signal, we must agree, is anything that produces a response, as just described, *without providing all of the power for the response*. The italics stipulate that at least some (often most) of the power for the response must come from the receiver itself. Anything that overpowers a receiver is not a signal. For instance, if one person sees danger approaching and pushes a friend out of the way, that is *not* communication. But if that person shouts, and the friend jumps aside, that *is* communication. The sound traveling through the air from the shouting person does not provide all of the power for the friend's response. The sound waves provide some power, because they must affect the friend's hearing, the neurons in the friend's ears for starters. The friend provides most of the power for moving itself.

The actions of a predator capturing its prey, or a parasite appropriating the metabolism of its host, are not signals. The predator or parasite provides most of the power for the capture or appropriation that changes the receiver. The interactions of predator and prey might include all sorts of communication, in the form of warnings and deceptions, but not the overpowering result. Likewise the interactions of parasite and host. Note that camouflage is a form of communication. Camouflage changes a predator's behavior to prey in comparison to its behavior otherwise. If a predator behaves differently to prey that are not camouflaged, then camouflage has communicated.

Learning consists of communication, provided demonstration or explanation by a teacher – the signaler – leads, more often than by chance, to some alteration in the student – the receiver. There is more to learning, though, as one of the following essays explores. Gossiping is also presumably communication, even when it only promotes friendship with little or no transfer of other information.

Chemical reactions, by themselves, are not communication, although one substance produces a change in another. All of the power required comes from the arrangement of molecules involved. Consider the case of hearing, just mentioned. A sound of course produces chemical changes in the neurons of a receiver's ear, but this is not all of the response. The entire response depends on the constitution of the hearer, in this case a nervous system and muscular system that result in movement by the hearer. We might view a living organism, such as this hearer, as a complex catalyst that chemically facilitates a particular form of jumping following a particular pattern of sound. Indeed this view is correct. It emphasizes that the receiver must provide some of the power, even in the form of chemical energy. It also emphasizes another universal feature of communication. Communication occurs between living organisms.

Living organisms are complex mechanisms for associating external events with internal changes. This complexity has accumulated gradually by means of evolution. Every living organism, even bacteria, single-celled organisms, and the simplest multi-celled organisms are mechanisms for associating a host of external situations with equally diverse responses, in most cases advantageous for the organism. Living organisms differ from inanimate ones in having evolved to respond appropriately to diverse situations.

In the presence of noise, these responses require decision-making. Each organism's sensors are its only input for decisions. A sensor might be a complex organ, such as an eye or an ear, or specialized cells, such as those for olfaction or touch. A sensor might be something as simple as a bacterium, with special chemical reactions asymmetrically distributed within its cell membrane. They might respond to external stimulation or to internal stimulation from various parts of its own body. Each time an organism checks its sensors, it might initiate a response to the presence of an appropriate stimulus. The response could affect its behavior immediately or it might be stored as a memory and perhaps affect behavior later. It might be immediate and overt, or it might be covert or delayed. Regardless of the nature of a sensor, noise changes everything.

Noise is stimulation that affects sensors in ways that can match the stimulation from a signal appropriate for a response.

All that an organism can detect is activity in its sensors. The same level of activity sometimes produced by appropriate stimulation, mixed with noise, is at other times produced by noise alone. For each sensor at each moment, an organism must choose some criterion for separating the presence of an appropriate signal, with noise, from the presence of noise alone. This criterion might be a simple threshold for activity in the sensor or it might be a complex criterion requiring advanced cognition.

Because of noise, all organisms, all of the time, must make decisions about whether or not to respond. Each time the organism knows *only two* states of its sensors. They either meet the criterion for a response to an appropriate stimulus, or they do not. Yet, in the presence of noise, each decision to respond or not has *four* possible outcomes – a response when an appropriate signal is actually present (a correct detection), a response when only noise is present (a false alarm), no response despite the presence of a signal (a missed detection), or no response when no signal is present (a correct rejection).

In two of these four cases, a false alarm or a missed detection, the receiver has made an error in detecting the appropriate signal. Noise creates this inescapable conundrum for every organism. It has long been recognized that living organisms are sensory and motor mechanisms. Because of noise, they are also all decision-making mechanisms.

Discussions of communication for the past century have proposed ways to classify signals (for instance, iconic or non-iconic) or to distinguish signals from other forms of stimulation. It is also sometimes proposed that signals are distinguished by a signaler's purposefulness or intentions, while cues are not. Any accidental or incidental stimulation from an organism is supposedly just a cue.

The following essays emphasize that this distinction is a spurious diversion. A signal is any pattern of matter or energy that produces a response by a receiver, without providing all of the power. Any purposefulness comes only from the evolutionary adaptation of the signal to produce an advantageous response as efficiently as possible. Any form of matter or energy can make a signal – sound, light, touch, combinations of molecules in odors or tastes. A signal can be continuous or intermittent, frequent or scarce, conspicuous or camouflaged, in any sensory modality. Pattern, some distinguishing arrangement of matter or energy, is important. Because a receiver's response is statistical, as emphasized above, communication occurs when a pattern of stimulation *on average* produces a response by a receiver.

This view of signals leads to clarification of what a signaler is. It is a living organism that has evolved to produce actions that can serve as signals to evoke advantageous responses in appropriate receivers. These actions are provoked by changes in its own constitution, in its own neural or other physiological states. These actions are often responses to signals that its own sensory organs have received. A bird, for instance, hears another individual's song and then responds with a song of its own, which can become a signal to the first or to others. All of these actions are embedded in noise, in the form of variability in the performance of an organism's own physiological systems as well as variability in the stimulation it receives.

A receiver in turn is a living organism that responds to signals, often signals mixed with noise. Its responses are embedded in noise, in the form of variability in its own physiological systems as well as variability in the sensations it receives. Receivers like signalers operate in a noisy environment with noisy physiological systems. The most important result of the mathematical description of signalers and receivers in the presence of noise is the impossibility of ever completely escaping noise. Diminishing benefits and escalating costs for an approach to perfection optimize behavior short of perfection.

Responses to signals require perception of signals by receivers. Note, however, that stimulation might come from the inanimate environment as well as from other organisms. In this case also, perception occurs. A receiver responds to a consistent pattern of stimulation in its inanimate environment, a pattern embedded in noise. As the following essays argue, communication and perception pose similar problems for a receiver. In both cases, noise makes any receiver a decision-making mechanism. Noise unites communication and perception.

Nevertheless, when signals originate with a signaler, it makes an enormous difference. Both receiver and signaler, as living organisms, evolve. The evolution of adaptations in communication require a compounded evolution. Both signalers and receivers, both production of signals and responses to signals, evolve jointly. Natural selection acts jointly on two organisms. It is their relationship that matters.

The evolution of adaptations for perception of inanimate situations requires no such complexity. Responses to temperature, daylight, salinity, overall terrain, currents, magnetic fields of the earth, are examples of responses that evolve by selection on single organisms acting alone. No doubt there are some situations in which a distinction between communication and perception fails (for instance, interactions between grossly different kinds of organisms), but there is no doubt that communication, as explored here, is a pervasive feature of living organisms.

To summarize, we can list these preliminary results, as follows:

1. Communication occurs when a living signaler produces a signal, to which a living receiver responds.

2. A response is a statistically predictable change in a receiver's state (including its nervous and other physiological system).

3. A signal is any pattern of stimulation from that produces a response by a receiver, without providing all of the power for the response.

4. In the presence of noise, any receiver must make decisions, with the possibility of some errors.

5. Communicating signalers and receivers evolve jointly.

The first chapter now begins with a review of the evolution of communication in animals, a field of study that began with Charles Darwin himself.

Chapter 1

Evolution of Communication

Introduction

Understanding the evolution of communication has undergone several saltations in the past century. Nonhuman animals are now routinely recognized to have spectacular and complex forms of communication. Also, after decades of controversy, it is now clear that communication is a form of cooperation. The conditions for the evolution of cooperation have also become clear. All of these statements can no doubt still excite controversy, but beyond any contention, they raise issues for the evolution of human language, as an extreme case of complexity in communication.

This article summarizes developments in evolutionary biology relevant to communication in general and introduces some implications for the specific case of language. Chapter 2, Features of Language, develops these implications in detail.

Comparative Study of Signaling

Study of communication by animals other than humans began in earnest with Darwin. Earlier concepts had placed organisms on an immutable *scala naturae*, with progressive elaboration of capabilities, including mental capabilities, from lower to higher, with a culmination among sublunary creatures in humans. Linnaeus' *Systema Naturae* (1735) instead formalized hierarchical classification of organisms, although humans still occupied first place with the unique attribute of wisdom.

Darwin first introduced natural selection of behavior in *On the Origin of Species* (1859), elaborated the possibilities in *The Descent of Man, and Selection in Relation to Sex* (1871), and illustrated applications to communication in *The Expression of the Emotions in Man and Animals* (1872). In subsequent decades, the study of animal behavior diverged into several paths: (1) experimental study of learning in a few convenient species, (2) experimental study of the sensory capabilities of animals, and (3) observational study of diverse organisms engaged in natural behavior.

The first path quickly established unsuspected capabilities for learning in animals and then investigated these abilities in species amenable to experimentation. The second path revealed that many animals had unsuspected sensory capabilities, including some unavailable to humans, such as complex vision, including ultraviolet and polarized light, ultrasonic sound, electric and magnetic fields, and echolocation. Mental capabilities of animals were no longer just subsets of human capabilities.

The third path, close observational study of animals, was at first pursued on the fringes of academic science. It eventually established at least six points important for a comparative study of communication.

(1) Many animals have sizable repertoires of actions, including vocalizations, not directly related to nutrition, survival, or procreation. Often they are relatively conspicuous, discrete, and stereotyped. The term "display" was appropriate for them.

(2) These displays are usually deployed in interactions between individuals and often evoke appropriate responses. They thus fit a basic criterion for communication. Furthermore, this communication is mostly among conspecific individuals. Both displays and responses are usually species-specific.

(3) The structural and behavioral traits of these displays reflect the phylogeny of species. Indeed, for a while it seemed that comparisons of these displays might reveal phylogeny better than morphology could. They appeared to have evolved in arbitrary directions, without the complications of convergent adaptations. Yet their stereotypy and elaboration (called ritualization) suggested adaptations for communication (Cullen 1966), and eventually it became clear that displays include many

adaptations to their environments and their functions in communication (Wilson 1965; Wiley and Richards 1982; Endler 1992).

(4) Detailed comparisons of behavior between and within species suggest that displays have often evolved by elaboration of much simpler actions, either actions for individual maintenance, incipient actions in other contexts, or actions that seemed partially inhibited or redirected in the circumstances (Tinbergen 1952, 1960)

(5) Experiments show that animals often respond only to a few simple features of displays (Tinbergen 1951). These “sign stimuli” often elicit relatively stereotyped responses, a finding that provided opportunities for a comparative neurobiology of behavior.

(6) The ontogenetic development of these displays and their corresponding responses often does not depend on shaping by reinforcement or encountering models. In other words, they are in many cases “innate” or relatively canalized, in the sense that they develop in a stable way despite normal variation in individuals’ experiences. In contrast, other actions and responses, equally complex and stereotyped, are learned by experience. In some cases, perhaps always, this learning is subject to constraints, predispositions such as sensitive periods, or templates. The first such case was imprinting of the following response by newly hatched precocial birds. Another was imitation of species-typical patterns of singing by songbirds. Predispositions in these cases are more canalized, within normal variation of experience, and simpler than the subsequent learned displays or responses. These examples of constrained learning have become epitomes of the interaction of genes and environment in the evolution of behavior, in particular, communicative behavior (Bateson 1981; Marler and Peters 1977; Marler 1990; Soha and Marler 2001a, b).

The discovery of such widespread and complex communicatory behavior in animals, generated by nonintuitive developmental processes and enmeshed in diverse social interactions, raised many questions about its evolution.

Evolution of Honesty

Until 50 or so years ago, the evolution of societies was explained by cooperation among individuals. Because cooperation is mutually beneficial, the action of natural selection in promoting cooperation seemed easily understood. This naïve attitude was overturned by George Williams' *Adaptation and Natural Selection* (1966) and Richard Dawkins' *The Selfish Gene* (1976). In the first place, not all individuals in an ostensibly cooperative society benefit equally. Individuals, for instance, might reduce their exposure to predators by herding, but those near the outside of a herd have more exposure than those in the center. Williams and Dawkins emphasized that, if differences in individuals' social behavior are associated with differences in the genes they carry, any allele (variants of genes) associated with behavior contributing to greater survival and reproduction spreads in a population, while others do not. Thus an explanation for the evolution of social behavior by natural selection requires an analysis of how each individual's behavior in social interactions affects its reproduction and survival.

This sort of argument provokes questions about how the behavior (or any trait) of an individual is related to its alleles. This basic process of behavioral ontogeny is revealed especially clearly in the studies of constrained learning in animals. An individual's development involves an interaction, in the statistical sense, of its genes and environment throughout the course of its life. As a result, genes do not determine anything about an individual's development, but they influence all of it. The same is true of the individual's environmental experiences. Natural selection results from differences in the reproduction and survival of individuals, whose traits are thus influenced more or less, in one way or another, by their alleles.

An early application of this principle was the evolution of polygynous mating systems, those in which most females mate with a few males. An argument that successful males benefit from multiple matings is insufficient without an explanation for how females benefit. The "polygyny threshold hypothesis" proposed that in habitats with high spatial variability (for instance, grasslands and marshes, where many nesting birds have polygynous mating systems), a female could compensate for reduced parental help from a polygynous mate provided that her mate's territory provided access to more food and safer nesting sites (Orians 1969; Searcy and Yasukawa 1995). This new approach to the evolution of social behavior set the stage for a reassessment of the evolution of communication.

Communication often involves individuals in asymmetrical roles, males enticing female mates, opponents in aggressive encounters, and competitors for food or space. Mutual advantages or cooperation in communication is less clear in cases like these, in which one individual might benefit by deceiving the other about its strength or suitability. Honesty, in contrast, would require a benefit for an individual responding to a signal as well as a benefit for the sender. Dawkins and Krebs (1978) suggested that signals are usually not honest. Instead they manipulate receivers for the signaler's advantage, despite the receiver's disadvantage. They deceive rather than inform receivers. Alternatively, Zahavi (1975, 1999) suggested that receivers avoid this problem by responding only to costly signals, because only costly signals are honest.

Zahavi's original proposal included two specific conditions for honesty: (1) costs of signals must be wanton (more than necessary); and (2) signals must have a form that impacts the attribute that is of interest to a receiver. The first of these conditions separates the costs of producing signals into a necessary component, which assures detection by a receiver, and an excessive or "wanton" component, which ensures honesty. The second condition requires that a signal interferes with its own meaning, in the sense that it must compromise the condition of the signaler that interests the receiver. For instance, a signal indicating efficiency in collecting food might partially compromise an ability to find food; or one indicating skill at avoiding predators would partially increase a signaler's vulnerability, for instance, by attracting a predator's attention or approaching and perhaps taunting predators. Such a signal would assure a receiver that the signaler was good enough at the particular task to overcome the handicap. For these reasons, Zahavi stipulated that the "wanton" costs of signals are handicaps. Handicaps thus became conditions for cooperative communication in which both signaler and receiver benefited.

This handicap principle became a central tenet of the study of animal communication, primarily as a result of mathematical demonstrations that honesty in signaling required costs for signals. Grafen (1990a, b) and Maynard Smith (1991) used different approaches to show that (1) the cost of an honest signal must exceed 0 as a general rule but (2) the cost to the signaler or the benefit to the receiver could equal 0 when signalers were genealogically related to receivers. Maynard Smith and Harper (2004) nevertheless concluded that the costs for honesty must exceed a cost necessary to avoid ambiguity in communication. Many studies in the past three decades have demonstrated that signals are usually honest (receivers' responses have benefits in terms of survival or reproduction or have some correlated effect) and that signals have costs related to survival or reproduction (Searcy and Nowicki 2005).

The handicap principle, however, is vitiated by two problems (Wiley 2015, 2017; see also Getty 1998; Számadó 2011). (1) The mathematical analyses make no distinction between necessary and excessive costs and in fact demonstrate only that honesty requires signals with costs >0. It is difficult, perhaps impossible, to imagine a signal that has no costs whatsoever, in terms of energy, risks, time, or lost opportunities, any of which would affect survival or reproduction. These analyses thus make no predictions about how much cost honesty requires. (2) An analysis of the optimization of communication in the presence of noise (Wiley 2015) shows that there is no distinction between costs of signals that reduce ambiguity and those that do not. During joint optimization of signalers and receivers in the presence of noise, all costs are incurred in reducing errors by receivers. Furthermore, a distinction between manipulation and information in communication is misleading, once information is more clearly defined, as proposed below.

More important, Grafen's calculations confirmed, although without much emphasis, that receivers must benefit from their responses to signals, at least on average. If receivers incur net costs for responding to a signal, then these responses do not evolve, and thus the signals do not either. The same conclusion applies to responses to signals from potential mates, although in this case, a receiver's benefits from choosing a mate can include genes that influence survival and reproduction of the receiver's offspring (Pomiankowski 1987).

The principal conclusion of these analyses is thus not that honest signals must have costs but that both signaler and receiver must benefit from the responses. It is not necessary that every instance of a response to a signal has benefits. Instead, responses to signals must have benefits, on average, either immediate or delayed, for both signaler and receiver. Communication is indeed a form of cooperation, in which both parties do better on average by communicating than they can otherwise.

Evolution of Cooperation

Cooperation begins by one individual helping another at some cost to itself. The first step is thus an act of altruism, one that benefits another at a cost to the actor, with benefit and cost ultimately in terms of survival and reproduction. Helping, including signaling, fits this pattern. Alleles cannot spread in a population unless the individuals that carry them survive and reproduce more effectively than others. Consequently, the challenge is to determine how alleles for helping can spread when helping individuals incur net disadvantages in survival or reproduction. It turns out that altruistic individuals, for which helping others decreases their own reproduction or survival, can persist in a population. Nevertheless, "altruistic" alleles, for which association with helping decreases their frequency, inevitably disappear from a population. To reconcile altruistic individuals with selfish alleles, two possibilities are now recognized: (1) helping genealogical relatives or (2) receiving compensating benefits in the future.

The first occurs, for instance, when individuals help to raise a relative's offspring while not themselves reproducing. The example of honeybees, which had perplexed Darwin, is such a case. Somewhat similar cases have now been studied in scores of birds and mammals as well as numerous social insects (Koenig and Dickinson 2004; Bourke 2011). In many cases the helpers (or workers) in fact reproduce to some extent either concurrently or later in life. Nevertheless, in some cases, such as honeybee workers helping queens to reproduce, the helping individuals almost never reproduce as much as the individuals they help. William Hamilton (1964, 1970) showed that alleles of individuals with lower chances of reproduction could nevertheless spread in a population provided these individuals helped close genealogical relatives. Close relatives can have a copy of any allele associated with a helper's behavior, as a result of their descent from a recent common ancestor. This "kin selection" is thus a special case of natural selection. If individuals sacrifice their lives to save the lives of more than two siblings (or more than eight cousins), any allele associated with this behavior would spread. The condition for the spread of an allele associated with helping is $C < rB$, where C is the cost in survival or reproduction for the actor, B is the benefit for the recipient, and r is the coefficient of genealogical relatedness of the actor to the recipient (one-half for a sibling, one-eighth for a cousin). Recent debate has clarified this possibility (Nowak et al. 2010; Abbot et al. 2011).

These ideas were quickly applied to mammals and especially to birds in which breeding pairs often have several helpers that feed or protect their young but do not (or are less likely to) reproduce. The evidence indicates that in most such cases, kin selection cannot provide a sufficient explanation for helping. Nevertheless, in the preponderance of cases, helpers are closely related to breeders. Kin selection in such cases contributes to the spread of alleles for helping even if it does not completely explain it. These principles apply to communication. We can expect individuals to accept uncompensated costs for signals or responses if the condition above is met.

The second case, when helpers receive later benefits, could apply to genealogical relatives and thus augment kin selection for alleles associated with helping. It might also apply to individuals without close genealogical relatedness. One possibility for later benefits is reciprocation: do unto your neighbor as you would have (or at least can expect) your neighbor to do unto you.

Reciprocation though has its complications. Just because an individual helps another does not assure that the recipient will return the favor. In addition to inevitable random contingencies, a population could plausibly include individuals with alleles associated with accepting help but never reciprocating, in other words defectors or cheaters. The possibility for cheaters in an otherwise cooperative population is pervasive. Close attention to cases of helping reveals that reciprocation is not a physical necessity, so there is always the possibility of a mutant allele that predisposes individuals to cheat by skipping reciprocation. In any instance of helping, the recipient is always possibly a cheater and the helper possibly a sucker.

In communication, the norm is honesty, but the possibility of exploitation, in other words cheating, is always present, both for signalers and receivers. On average a signaler benefits from responses by appropriate receivers, but there is always the chance of an inappropriate receiver, such as an eavesdropping rival, predator, or parasite. These inappropriate receivers exploit the signals intended for appropriate receivers. This situation is the converse of deception. In deception, an inappropriate signaler exploits the responses intended for appropriate signalers. There is always the possibility that a signaler or a receiver is a cheater in normally honest communication, and there is always the possibility that a signaler or a receiver is a sucker.

The usual recourse for analyzing the evolution of reciprocation has been game theory. In particular, the prisoner's dilemma and many related games have provided a foundation for mathematical and experimental analysis of the evolution of cooperation. An early discovery was that the behavioral strategy of iterated tit for tat permits the evolution of cooperation by reciprocation (Axelrod and Hamilton 1981; Axelrod 1984). This strategy consists of helping any new partner on the first encounter and then on subsequent encounters, either helping or not depending on whether or not the partner has reciprocated (Axelrod and Hamilton 1981; Axelrod 1984). It is a practical variant of the golden rule: do unto each partner as you expect that partner to do unto you.

Many additional possibilities for the evolution of altruistic behavior have by now surfaced. For instance, altruistic behavior can evolve in local populations of sessile organisms, when individuals interact repeatedly with the same few partners as a result of their immobility. With the strategy win-stay-lose-shift, individuals help in their first interaction (or in random occasional interactions) and then help subsequent partners or not, regardless of whom they might be, depending on whether or not the previous partner reciprocated. In this case individuals do not keep record of their partners. Individuals might also develop a positive reputation for helping so that others would help them in the expectation that they would receive help in return. In this case, however, an individual's tendency to help would have to extend indiscriminately to other individuals. Alternatively individuals might develop a negative reputation so that others would have no expectation of reciprocation and thus refuse to help them. Finally, helpers might punish (impose extra cost on) defectors. In this case, helpers would themselves incur an additional cost for administering punishment. If the cost to each punisher was sufficiently small (perhaps shared among many helpers) and the cost to each punished non-reciprocator was sufficiently large (perhaps execution or banishment with little chance of survival or reproduction), this possibility could overturn the advantages of cheating but preserve most of the advantages of helping. Sharing the costs of punishment would, however, be another form of reciprocation, which would itself open opportunities for cheating (for instance, avoiding a fair share of taxes for policing). Sometimes cases of social approval or social disapproval are combined as contrasting examples of "social selection." All of these possibilities, when appropriate conditions are met, can explain how alleles associated with altruistic action can persist or spread in a population (Nowak 2006; for an example of punishment by monkeys, Hauser 1992).

There is one situation that does not allow such altruistic alleles to spread – badges identifying altruists. Altruistic individuals could increase the chances of reciprocation by recognizing each other by some badge associated with helping.

Richard Dawkins labeled this possibility a "green-beard effect" (on the possibility that green beards might serve for such a badge). The problem is that a shared badge just creates another opening for cheaters, individuals that sport the badge but do not reciprocate. Alleles with pleiotropic effects might work temporarily but only until a mutation broke the association between helping and development of a badge.

From decades of discussion, two conditions have emerged that always increase chances for the evolution of mutual helping: (1) genealogical relatedness between actors and recipients and (2) cognitive abilities for remembering interactions with individual opponents. Genealogical relatedness promotes the evolution of cooperation even when it does not provide a complete explanation. Memory of individual opponents is a crucial part of some options for the evolution of cooperation by reciprocity, such as tit for tat. Some options, such as tracking individuals' reputations, require extensive memory of individuals. Nevertheless, even possibilities that require no memory of opponents, such as sessile neighborhoods or win-stay-lose shift, are even more likely to evolve reciprocity when memory of individual opponents is possible.

These possibilities for multiple contributing factors in the evolution of cooperation are often overlooked, because theoretical models have tended to take two directions, either focusing on the minimal conditions or on the maximal potential for cooperation. Those studying humans focus on the maximal possibilities for cooperation. Those studying other animals tend to focus on the minimal requirements for cooperation. Studies with humans in mind often assume complex cognitive capabilities without much comment and ignore the synergistic contribution of genealogical relatedness. They also can ignore the possibility that collaboration in a complex society, which results in unequal or uncertain advantages for individuals, generates natural selection for defection and for nothing-to-lose retaliation when disadvantaged. All these conditions for the evolution of cooperation apply to the evolution of communication.

Evolution of Individual Recognition

For the evolution of cooperation, all but the simplest possibilities rely on individuals' abilities to remember other individuals and to associate them with particular patterns of behavior. This ability requires at least minimal object constancy for another individual, discrimination of that individual from others, and association of that individual in memory with its previous behavior. Object constancy, discrimination, and association are mental processes, perhaps each some aspect of association in general, that recur in all discrimination learning.

Abilities of this sort are now well documented for many nonhuman organisms. There are important distinctions to be made about the complexity of recognition (Wiley 2013a). First, the specificity of individual recognition can vary. An experiment might reveal that subjects respond to a neighbor or a partner (or often just to some features of such an individual) in a different way than to other individuals. Such a discrimination could result solely from habituation to repeated experience with a familiar individual. It would thus constitute recognition of a particular individual only when no other individual could have such familiarity. In some cases it is still not clear whether or not animals recognize territorial neighbors, parents, offspring, or relatives as individuals or as small sets of familiar individuals and whether or not these small sets are distinguished from others by associative learning or solely by habituation. On the other hand, experiments have shown that many territorial birds, for instance, can identify specific individuals within the small sets of their familiar neighbors.

Specificity of individual recognition is crucial for the evolution of cooperation. Only when specific individuals are recognized can interactions with possible cheaters be avoided or reduced. One report of a nonhuman animal that fits the requirements for tit for tat provides an example. A warbler in eastern North American forests can use its ability to recognize individual territorial neighbors in order to cooperate with them once mutual boundaries are settled. Experiments with playbacks of songs show that they use tit for tat to retaliate specifically against defecting (trespassing) neighbors (Godard 1993).

The multiplicity of individual recognition can also vary, from recognition of a single other individual (for instance, a mate) to recognition of individuals in a small set (perhaps territorial neighbors, a small group, or current offspring) to recognition of potentially large numbers of individuals (as do humans). Some territorial birds are known to recognize several individual territorial neighbors, and primates (and presumably some other social birds and mammals) can recognize multiple individuals within their social groups and in nearby groups as well (Cheney and Seyfarth 1990, 2007). Associations with these individuals are probably not complicated in the case of territorial neighbors in particular locations but are possibly more complex in the case of group members encountered in diverse contexts.

Humans recognize large numbers of individuals with different degrees of specificity and different complexities of associations. There is a considerable literature on the influence of features, relationships of features, and contexts in peoples' abilities to recognize faces, but almost nothing is known about how many individuals a person can recognize nor about the cognitive complexities of how a person organizes these memories. This gap in our knowledge is surprising, because the memory, associations, categories, and relationships involved in individual recognition by humans seem to approach those needed for human language.

The hierarchical organization of many animal societies suggests possibilities for recognition that have parallels with language. The formation of a dominance hierarchy, it is important to acknowledge in the first place, might not require any individual recognition at all. Each individual can plausibly learn to recognize sets of higher- and lower-ranking opponents or even respond to one or more graded features (for instance, size, postures, or badges of dominance). On the other hand, individuals might recognize each individual opponent and respond to each in a different way. Associations with each opponent might allow inferences about the relative ranks of any two other individuals.

Especially interesting are cases in which individuals' rankings include prominent subgroupings, and ranks across these subgroups do not follow gradations of individuals' features. In many primate groups, for instance, matrilineal groups as a whole are ranked as well as individuals within each matriline (Bergman et al. 2003). Thus a low-ranking young individual in a high-ranking matriline outranks a high-ranking older female in a low-ranking matriline. These nested hierarchies seem to arise because older relatives (mothers, aunts) shield younger ones from subordination by individuals in lower-ranking matrilines. Although the pattern and the interactions are well documented, there remains the question whether this pattern is conceptualized by individual monkeys as embedded subgroups or as an overall hierarchy. Some evidence suggests that the former is possible for baboons. Playbacks of calls indicating a reversal of ranks between matrilines evoke more attention than those indicating a reversal within matrilines, regardless of the differences in overall rankings (Bergman et al. 2003). Baboons evidently can conceptualize a dominance hierarchy as sets of embedded individuals, although wide overlap in the ranges of responses to the two conditions raises the possibility of inconsistency in this ability.

A similar situation results from "coattail" effects in dominance hierarchies of birds (Wiley 1990; Cristal 1995). Small groups of emberizine sparrows are allowed to form dominance hierarchies in large cages in winter, when competition for food is the predominant activity. Then the top half of the hierarchy in one cage and the top half from a second cage are combined in a third neutral cage. Surprisingly, the two groups often remain coherent in the newly formed hierarchy. Each individual's rank is nested in its group's rank. The mechanism is perhaps not dissimilar to that in primate groups. The two highest-ranking individuals in a combined group interact to determine their relationship, but once this relationship is decided, the higher individual creates a coattail for its familiar subordinates. In this case, and perhaps in the primate groups also, the principal effect of the dominant individual is to let its familiar opponents approach more closely than can others without aggression. Again the question arises: How does a relatively dominant member of a subordinate subgroup categorize opponents? First by subgroup and then by ranking within it or simply by overall rank?

Evolution of Mating Preferences by Sexual Selection

Mate choice is a well-studied example of communication. In many animals males perform conspicuous displays that increase their chances of mating. Darwin (1859, 1871) recognized that if females prefer males with certain traits, or if males with certain traits are more successful in competing with rival males, then these traits would tend to spread in a population. Even more than his theory of natural selection, this theory of sexual selection precipitated controversy among biologists. At first the primary sticking points were doubts about the cognitive abilities of females needed for preferences, but R. A. Fisher (1930) made it clear that a female's preference is no more than a neurophysiological response to a male's traits. Fisher then presented a verbal argument for accelerating evolution of male traits preferred by females. The process of sexual selection was terminated when the cost of the preferred male traits became too great.

It remained for proper mathematical analyses to verify the dynamics of this accelerating evolution (Lande 1981; Kirkpatrick 1982). Much subsequent work has confirmed predictions about mate choice in natural populations (Anderson 1994; Searcy and Yasukawa 1995). The key to the evolutionary dynamics is the genetic correlation produced when a female with a preference mates with a male with a preferred trait. Their offspring tend to inherit alleles associated with both preference and trait. The result is a genetic correlation between these alleles in the population. Within the genomes of individuals in the population, the presence of the preference allele is associated with the presence of the trait allele. This association is often called "linkage disequilibrium" by geneticists, but linkage is actually a special case of genetic correlation, not necessarily connected with mating preferences. As generations pass, females with preferences tend to spread not only alleles for the preferred male trait but also alleles for the preference (because both males and females tend to carry both alleles). As a result of the genetic correlation of the two alleles, the preference allele spreads by "hitchhiking" with the trait allele. Another way to look at it, the preference allele spreads itself. The genetic correlation that results from preferential mating produces accelerating evolution of alleles for both of the male trait and the female preference. Sexual selection is thus a special case of natural selection, one that happens whenever individuals of one sex with a particular trait mate disproportionately with members of the other sex with the same or different trait. Preferences and other traits, to reiterate a point above, develop under the influence of alleles.

There are several points to emphasize here. First, A preference for potential mates with a particular trait is a form of communication. In the most frequent case, males produce signals to which females respond discriminatingly. Nevertheless, the mathematical models of sexual selection do not require direct choice of males' traits. A female might instead exert a choice indirectly. She might set conditions for mating by provoking a contest between potential mates. For instance, she might limit her matings to a particular time and place, or she might indiscriminately advertise her readiness to mate. In these cases there is no discrimination between males' traits, yet females set conditions that result in selective mating. By mating with whichever male prevails in such contests, she would indirectly choose a male whose traits allowed him to prevail against all comers. Females would, in other words, define the contest for males and then take any winner as a mate. In the case of indiscriminate advertising, a female produces a signal to which males respond by approaching. In any of these possibilities for indirect choice, genetic correlation and subsequent sexual selection would result, just as in the case of direct choice.

Second, Alleles for preferences cannot spread if females with these alleles incur net costs in terms of survival and reproduction (Pomiankowski 1987; Grafen 1990b). Responses to signals, as we saw above, must result in net advantages, on average, for receivers or their close relatives. Furthermore, sexual selection also stops when males incur net costs, when the advantages of greater possibilities for reproduction are more than offset by disadvantages for survival.

Third, Sexual selection does not spread alleles until the frequencies of the preference and trait alleles in the population exceed a threshold (or the level of genetic correlation crosses a threshold) (Lande 1981; Kirkpatrick 1982). Sexual selection does not spread mutations ab initio.

This hurdle applies to the initial evolution of any signal and response. No matter how advantageous communication might be, neither a response nor a signal can spread by itself. A rare mutant for a new response cannot spread without sufficiently frequent signals, and vice versa a new signal cannot spread without sufficiently frequent responses. Mutualistic signal and response must overcome a hurdle before they can spread.

Furthermore, all mutualistic interactions spread in an accelerating way once started. Responses become more advantageous as the frequency of signals increases and vice versa. It is still not clear whether R. A. Fisher (1930) had genetic correlation in mind when he proposed accelerating evolution of sexually selected traits or whether he was just thinking of the accelerating spread of any frequency-dependent mutualistic interaction (Wiley 2015). The rate of spread eventually slows down as the frequencies approach fixation, because increasing frequency of signalers results in diminishing advantages for receivers and vice versa.

Sexual selection, despite its specific application to communication during mate choice, includes parallels with the frequency-dependent evolutionary dynamics of all forms of mutualistic interaction and thus of communication in general. The evolution of mutualistic communication between individuals other than mates does not receive the extra boost from the genetic correlation that results from mating. Nevertheless it does share the initial hurdle and the subsequent acceleration that apply to all frequency-dependent mutualism.

A final point should be emphasized. The original mathematical models (Lande 1981; Kirkpatrick 1982) and subsequent derivations include the possibility that female preferences might be arbitrary. Arbitrary in this context means that mating with a preferred mate provides no benefit whatever to the female (we have already emphasized that it cannot incur a cost to the female). There are three reasons to think that such completely arbitrary preferences are unlikely to evolve. First, if alleles for two preferences exist in the same population, the one that results in a greater benefit to females spreads faster. So any preference with a benefit for females spreads to the exclusion of an arbitrary preference. Second, the same applies to the costs of male traits. Of two alleles associated with traits equally preferred by females, those with lower costs spread fastest. Finally, preferences for traits are a form of communication. Noise in communication makes the evolution of completely arbitrary signals and responses unlikely. The parameters, some ten of them, that influence the utility of a signal for a signaler (an advertising male, for instance) and the utility of a criterion of response for a receiver (a choosy female, for instance) would have to balance exactly to produce zero net utility for both signaler and receiver (Wiley 2015, 2017). Communication for mate choice, like all communication, is inescapably noisy.

Evolution of Communication in Noise

Noise requires a new approach to understanding the evolution of all forms of communication, one that is compatible in part with the preceding approaches but has advantages of defining some crucial concepts, presenting a thorough optimization of the behavior of signalers and receivers and incorporating the consequences of noise for communication. Noise opens a new perspective on the evolution of communication. Most surprising is the realization that evolution is not expected to produce noise-free communication.

Everyone is aware that noise can interfere with communication. Communication requires two parties, a signaler and a receiver. Even when more than one signaler or receiver is active at the same time, each instance of communication is a relationship between a signaler and a receiver. A signaler produces a signal to which a receiver might respond. A signal is any pattern of energy or matter that can elicit a response from a receiver, without providing all of the power for the response. A response need not occur every time a signal is perceived, but unless a response occurs more often than at random, there is no evidence for communication.

Most previous definitions of a signal agree in stipulating that a signal must evoke a response, although it is less often emphasized that responses need to occur only more often than random. Previous definitions also require that a signal must have evolved for the purpose of communication or have a goal (or intention or function) of evoking a particular response. Often there is a complementary condition: a response must have evolved for a particular signal. These stipulations are confusing and circular: signals and responses evolve for communication which consists of signals and responses.

To avoid this confusion, a signal is defined as a pattern of energy that evokes a response but does not provide all the power for the response. The restriction on the power of a signal excludes cases in which one individual simply overpowers another, as, for instance, in predation. A signal, on the other hand, must provide some power, enough to activate the sensory receptors of a receiver. The receiver must then provide some, often most, of the power for the response. Consequently, receivers have the final control of responses.

Signals defined in this way can originate from inanimate objects as well as living ones. Previously most definitions have excluded this possibility by insisting that signals have functions; signals without functions are instead called cues. But this measure is unnecessary. From a receiver's point of view, it makes no difference what the source of a sensation is. Signals from any source produce sensations for receivers, and receivers have final control of responses. For any receiver, including humans, information about the inanimate world has the same footing as information about the behavior of other organisms.

Nevertheless, signaling and responding by organisms can evolve. As discussed above, signals by evolving signalers are expected on average to produce net advantages in terms of survival or reproduction for signalers; responses produce such advantages for receivers. Signaling and responding by living organisms evolve jointly.

Information is another term that has caused confusion. Shannon in his pioneering papers on information theory (Shannon and Waver 1949) described the simplest intuitive way to measure the quantity of information, but he only hinted that the quality of information ("what" rather than "how much" information) depended on the state of the signaler. A signaler's state results from its ontogeny, as described above, the accumulated influences of genes and experiences during the course of its life to present. Its state is the current condition of its body, including its nervous system and thus also its recent perceptions. The quality of information, regardless of its quantity, is the correlation between a signal and the signaler's state (Wiley 2013b). If a signal has information relevant to a receiver's survival or reproduction, alleles associated with responding appropriately to such a signal can evolve (increase or decrease in frequency in the population). If such a signal evokes a response affecting the signaler's survival or reproduction, alleles associated with producing such signals can evolve. Thus the quality as well as the quantity of information in a signal influences its evolution. Evolving signals must include some information about the signaler's state. Note once again that responses can include delayed and covert effects, such as memory or physiological changes affecting later behavior. It might include complex perceptions as well as simple reflexes.

With collateral issues resolved, a criterion for noise is possible: noise is errors by receivers. This insight by Shannon is as important as that about information. Anything that results in errors by receivers counts as noise. It can include irrelevant background energy that interferes with a receiver's detection or discrimination of signals. Background energy can include turbulence and extraneous energy impinging on the receiver's sensors. It can include signals of other species or individuals irrelevant to the receiver in question.

Noise can result from attenuation and degradation of the patterns of signals during transmission from signaler to receiver, in the atmosphere or water or even at the interface between a finger or a tongue and a receiver's skin. Noise also occurs in nervous systems. Nearly all neurons produce action potentials continuously at irregular rates, which are combined with the firing rates elicited by sensory stimulation. Little is known about how this neural noise affects perceptions of sensory sensations. Both signalers and receivers are subject to neural noise. Signalers do not always produce signals perfectly correlated with their states; receivers do not always respond appropriately to signals (Wiley 2015, 2017).

The basic insight of signal detection is that a receiver must make a decision each time it checks any of its sensory inputs (Macmillan and Creelman 1991, 2005). Because relevant signals combine with noise in a receiver's sensors, a receiver must decide whether a sensation correctly indicates the occurrence of a signal or not. Normally the combination of signal plus noise results in greater stimulation than does noise alone. Both signals and noise vary, so each produces a probability density function

(PDF) of levels of excitation of a sensor. Whenever the PDF for signal plus noise and the PDF for noise alone overlap to any extent, the receiver cannot respond to appropriate signals without some error.

Another insight of signal detection is that receivers are subject to two kinds of errors, false alarms and missed detections (or errors of commission or omission or Type I and II errors in statistical comparisons). These errors result from a receiver's criterion for a decision to respond or not. The simplest criterion for response is a threshold. If the level of excitation in a sensor exceeds the threshold, then respond; otherwise, do not. More complex sensors can include filters and combinations of thresholds and filters to produce specific cognitive criteria for a response. A response, as emphasized above, can be either overt or covert, an act, a perception, or a memory. Any response might be an error.

Every time a receiver checks its input, exactly four mutually exclusive outcomes are possible. The level of excitation in its sensor might exceed its threshold (or other criterion for response) or not; in each case, a relevant signal might have occurred or not. Excitation above threshold when a signal is present results in a correct detection. With no signal, only noise, the result is a false alarm. Excitation below threshold when only noise is present results in a correct rejection; with a signal present, albeit attenuated and masked with noise, the result is a missed detection.

A receiver in this situation can adjust its rates of error by adjusting the level of its threshold. Yet every adjustment of a threshold changes the probabilities of all four possible outcomes. For instance, by raising its threshold for a response, it would reduce its chances of a false alarm. Its chances of correct detection also decrease. On the other hand, missed detections and correct rejections increase. If a receiver lowers its threshold, its chances of a missed detection decrease, but again the probabilities of all other outcomes would also change. Most important, the two kinds of error, false alarm and missed detection, always change in contrary ways. A receiver cannot reduce one kind of error without increasing the other. Every time a receiver checks its sensors and decides to respond or not, it is in an inescapable double bind.

Signalers can influence the relationship between signals and noise for receivers. In general the more powerful or concentrated a signal, the greater its impact on the sensors of a receiver at any particular distance. Signalers thus can increase the probability of a correct detection (an appropriate response) by a receiver by increasing the exaggeration of a signal. The probability of a correct detection by an appropriate receiver in turn affects the benefit from signaling for a signaler.

The next objective is to calculate how natural selection can affect the evolution of signalers and receivers in noise. Does evolution by natural selection produce a joint solution for signaling and responding? Are there mutual advantages for signalers and receivers? Is noise eliminated by natural selection on communication? The first step is to specify net advantages and disadvantages for potential options for both receivers and signalers (Wiley 2015). For receivers, this step requires calculation of the utility of the receiver's threshold, each time a receiver checks its sensors and decides to respond or not. This utility is a function of the signal/noise ratio (more accurately, the relationship of the PDFs for signals plus noise and for noise alone), the probability of a signal, and the level of the threshold. The threshold affects the probabilities of each of the four possible outcomes. For investigating evolution, the utility is expressed in terms of the receiver's survival reproduction.

A similar process can specify the utility of signal exaggeration for a signaler as a function of the cost of producing a signal with a particular exaggeration and the probability of a response from an appropriate receiver. The utility of a receiver's threshold thus depends on the signal/noise ratio, which depends on the exaggeration of a signaler's signal; conversely, the utility of a signaler's exaggeration depends on the probability of a correct detection, which depends on the level of a receiver's threshold. With some calculus, it is possible to find the optimal level of a threshold for a given exaggeration of a signal or, alternatively, the optimal exaggeration of a signal for a given threshold. Further computation reveals that in most situations, thresholds and exaggeration evolve by natural selection to a joint optimum, a Nash equilibrium at which both receivers and signalers have advantages and both do as well as possible provided the other does also (Wiley 2015).

General Principles for the Evolution of Communication

Several important insights about communication result from these calculations. First of all, adaptation by natural selection does not escape from noise. Noise is inevitable; communication in the absence of noise is unattainable. This conclusion follows from the basic insight that both signalers and receivers face conflicting advantages and disadvantages. They also face diminishing returns: in an approach to optimal signals or thresholds, advantages increase less rapidly than disadvantages.

Second, the variables in these calculations are completely general. They apply to signalers and receivers in all cases of communication. For instance, both exaggerations of signals and thresholds for response scale to the level of noise. Thus at close range, when background noise is minimal, optimal signals and thresholds decrease in relation to noise so that possibilities for error persist. In addition, this result applies not only to communication by nonhuman animals but also to humans and to all modes of human communication including electronic. It applies not only to communication between organisms but also to communication between and within cells. Molecular signals and receptors operate in a noisy environment, with multiple signals, multiple receptors, and chemical degradation. There is thus no reason to expect that adaptation by natural selection can eliminate noise in any form of communication. The same conclusion applies to perceptions of the external world. Optimal decisions by perceivers, based on signals from inanimate objects, cannot escape some probability of error, either false alarms or missed detections and errors of commission or omission.

Third, these calculations confirm the results of all previous calculations of the evolution of mutual interactions: neither signals nor responses can spread when both are infrequent. In the present calculations, when thresholds are too high (as when individuals have little tendency to respond to a particular stimulus) and when exaggeration of signals is too low (as when individuals have little tendency to produce them), communication collapses. Thresholds for response must not be too high, and exaggeration of signals must not be too low, for mutual communication to evolve ab initio.

Fourth, optimization of communication in noise opens many questions about adaptations for communication in different circumstances. Nearly every parameter in the utility functions has been measured or estimated in some case of communication (Wiley 2015), although never has there been a complete analysis of communication in noise in natural situations. There are clear predictions about the evolution of thresholds for response and for exaggeration of signals as a result of differences in signal frequency and the utilities of the four possible outcomes for receivers. A further prediction is that exaggeration of signals should not evolve in arbitrary ways but instead specifically in ways that increase the signal/noise ratio for appropriate receivers. There is evidence that evolution of both acoustic and visual signals follows this prediction (Wiley 1991; Endler 1992; Endler and Thiéry 1996; Gomez and Théry 2004; Kemp et al. 2009). It also seems unlikely that purely arbitrary signals could ever evolve even by sexual selection. An arbitrary signal would require the multiple parameters of communication in noise to balance exactly.

Finally, although communication is never expected to reach perfection, honesty is expected to prevail. At the evolutionary optimum, both signalers and receivers benefit on average. Nevertheless, some incidence of error persists. Some errors result from random events in the environment or in signalers or receivers. Evolved adaptations by inappropriate participants also affect optimal behavior by signalers and receivers.

Deception by inappropriate signalers can reduce the utility for receivers. Eavesdropping by inappropriate receivers can reduce the utility for signalers. Communication is expected to transfer information between signaler and receiver, in other words, a corresponding perception of the actual world to their mutual advantage but always with a possibility for random errors and for manipulation by inappropriate signalers or receivers.

Investigation of communication with a hypothetical absence of errors is thus unrealistic. Noise requires study as much as signals. Nevertheless, many studies of communication, whether theoretical, observational, or experimental, take steps to reduce noise as much as possible, in order to focus on signals or responses. Study of simplified situations has its place in science, but it can also produce unrealistic results. To understand communication in any real situation, noise is as important as the signals. In other words, communication of any sort cannot be fully understood without understanding its variation in practice.

It is also clear that all communication in noise is “inferential” and “intentional.” If the “meaning” of a signal is the response it evokes in a receiver (whether overt or covert), the meaning in the presence of noise always requires a decision by a receiver. Meaning is thus “inferential.”

Furthermore, signals are often accompanied by relevant as well as noisy contexts, both of which affect the receiver’s decisions to response or not. These decisions must often depend on the receiver’s previous interactions in a particular context and thus on the receiver’s memory of any associations with this context. They thus are “intentional.”

Signals always arrive within a context of noise. A receiver decides to respond based on its current state (including memory) and the sensations it receives. These sensations include signals (usually honest but with some noise) and relevant contexts (usually correct but with some noise). If “inferential” implies decisions by a receiver and if “intentional” implies associations for a receiver between a signal and its context, then all communication is intentional, between animals as well as humans. The decisions humans make in using language require complex criteria for responses. The distinctive features of human linguistic communication as opposed to other forms of communications lie in the specific complexities of these cognitive criteria, not in the importance of decisions or contexts in general.

References

Abbot, P., Abe, J., Alcock, J., Alizon, S., Alpedrinha, J. A. C., Andersson, M., Andre, J.-B., et al. (2011). Inclusive fitness theory and eusociality: Reply to Nowak et al. *Nature*, 471, E1–E4. https://doi.org/10.1038/nature09831.

Andersson, M. (1994). *Sexual Selection*. Princeton: Princeton University Press.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

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

Bateson, P. (1981). Control of sensitivity to the environment during development. In K. Immelmann, G. W. Barlow, L. Petrinovich, & M. Main (Eds.), *Behavioral Development* (pp. 432–453). Cambridge, UK: Cambridge University Press.

Bergman, T. J., Beehner, J. C., Cheney, D. L., & Seyfarth, R. M. (2003). Hierarchical classification by rank and kinship in baboons. *Science*, 302, 1234–1236.

Bourke, A. F. (2011). *Principles of Social Evolution*. Oxford: Oxford University Press.

Cheney, D. L., & Seyfarth, R. M. (1990). *How Monkeys See The World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University of Chicago Press.

Cristol, D. A. (1995). The coat-tail effect in merged flocks of dark-eyed juncos: Social status depends on familiarity. *Animal Behaviuor*, 50, 151–159.

Cullen, J. M. (1966). Ritualization of animal activities in relation to phylogeny, speciation and ecology: Reduction of ambiguity through ritualization. *Philosophical Transactions of the Royal Society, London*, B, 251, 363–374.

Darwin, C. (1859). *On the Origin of Species*. London: John Murray.

Darwin, C. (1871). *The Descent of Man and Selection in Relation to Sex*. London: John Murray.

Darwin, C. (1872). *The Expression of the Emotions in Man and Animals*. London: John Murray.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dawkins, R., & Krebs, J. R. (1978). Animal signals: information or manipulation. In J. R. Krebs and N. B. Davies (Eds.), *Behavioral Ecology: An Evolutionary Approach* (Ed. 2, pp. 282–309). Oxford: Blackwell.

Endler, J. A. (1992). Signals, signal conditions and the direction of evolution. *American Naturalist*, 139 (Supplement), S125–S153.

Endler, J. A., & Thery, M. (1996). Interacting effects of lek placement, display behavior, ambient light, and color patterns in three neotropical forest-dwelling birds. *American Naturalist*, 148, 421–452.

Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Oxford University Press.

Getty, T. (1998). Handicap signaling: When fecundity and viability do not add up. *Animal Behaviour*, 56, 127–130.

Godard, R. (1993). Tit for tat among neighboring hooded warblers. *Behavioral Ecology and Sociobiology*, 33, 45–50.

Gomez, D., & Théry, M. (2004). Influence of ambient light on the evolution of color signals: Comparative analysis of a Neotropical rainforest bird community. *Ecology Letters*, 7, 279–284.

Grafen, A. (1990a). Biological signals as handicaps. *Journal of Theoretical Biology*, 144, 517–546.

Grafen, A. (1990b). Sexual selection unhandicapped by the Fisher process. *Journal of Theoretical* Biology, 144, 473–516.

Hamilton, W. D. (1964). The genetical evolution of social behavior, I & II. *Journal of Theoretical Biology*, 7, 1–52.

Hamilton, W. D. (1970). Selfish and spiteful behavior in an evolutionary model. Nature, 228, 1218–1220. Hauser, M. D. (1992). Costs of deception: Cheaters are punished in rhesus monkeys (Macaca mulatta). *Proceedings of the National Academy of Sciences*, 89, 12137–12139.

Kemp, D. J., Reznick, D. N., Grether, G. F., & Endler, J. A. (2009). Predicting the direction of ornament evolution in Trinidadian guppies (Poecilia reticulata). *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20091226. https://doi.org/10.1098/ rspb.2009.1226.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, 36, 1–12.

Koenig, W., & Dickinson, J. (2004). *Ecology and Evolution of Cooperative Breeding in Birds*. Cambridge, UK: Cambridge University Press.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78, 3721–3725.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge, UK: Cambridge University Press. Second edition (2005), Mahwah, NJ: Lawrence Erlbaum.

Marler, P. (1990). Innate learning preferences: Signals for communication. *Developmental Psychobiology*, 23, 557–568.

Marler, P., & Peters, S. (1977). Selective vocal learning in a sparrow. *Science*, 198, 519–521.

Maynard Smith, J. (1991). Honest signaling: The Philip Sidney game. *Animal Behaviour*, 42, 1034–1035.

Maynard Smith, J., & Harper, D. (2004). *Animal Signals*. Oxford: Oxford University Press.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314, 1560–1563.

Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature*, 466, 1057–1062. https://doi.org/10.1038/nature09205

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

Pomiankowski, A. (1987). The costs of choice in sexual selection. *Journal of Theoretical Biology*, 128, 195–218.

Searcy, W. A., & Nowicki, S. (2005). *The Evolution of Animal Communication.* Princeton: Princeton University Press.

Searcy, W. A., & Yasukawa, K. (1995). *Polygyny and Sexual Selection in Red-Winged Blackbirds*. Princeton: Princeton University Press.

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Soha, J. A., & Marler, P. (2001a). Cues for early discrimination of conspecific song in the white-crowned sparrow (*Zonotrichia leucophrys*). *Ethology*, 107, 813–826.

Soha, J. A., & Marler, P. (2001b). Vocal syntax development in the white-crowned sparrow (*Zonotrichia leucophrys*). *Journal of Comparative Psychology*, 115, 172–180.

Számadó, S. (2011). The cost of honesty and the fallacy of the handicap principle. *Animal Behaviour*, 81, 3–10.

Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Oxford University Press.

Tinbergen, N. (1952). "Derived" activities; their causation, biological significance, origin, and emancipation during evolution. *Quarterly Review of Biology*, 27, 1–32.

Tinbergen, N. (1960). Comparative studies of the behavior of gulls (Laridae): A progress report. *Behaviour*, 15, 1–69.

Wiley, R. H. (1990). Prior-residence and coat-tail effects in dominance relationships of male dark-eyed juncos, *Junco hyemalis*. *Animal Behaviour*, 40, 587–596.

Wiley, R. H. (1991). Associations of song properties with habitats for territorial oscine birds of eastern North America. *American Naturalist*, 138, 973–993.

Wiley, R. H. (2013a). Specificity and multiplicity in the recognition of individuals: Implications for the evolution of social behavior. *Biological Reviews*, 88, 179–195.

Wiley, R. H. (2013b). Communication as a transfer of information: Measurement, mechanism, and meaning. In U. Stegmann (Ed.), *Animal Communication Theory: Information and Influence* (pp. 113–129). Cambridge, UK: Cambridge University Press.

Wiley, R. H. (2015). *Noise Matters: The Evolution of Communication*. Cambridge, MA: Harvard University Press.

Wiley, R. H. (2017). How noise determines the evolution of communication. *Animal Behaviour,* 124, 307–313.

Wiley, R. H., & Richards, D. G. (1982). Adaptations for acoustic communication in birds: Sound propagation and signal detection. In D. E. Kroodsma & E. H. Miller (Eds.), *Acoustic Communication in Birds* (Vol. 1, pp. 131–181). New York: Academic.

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton: Princeton University Press.

Wilson, E. O. (1965). Chemical communication in social insects. *Science*, 149, 1064–1071.

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

Zahavi, A. (1999). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford: Oxford University Press.

Chapter 2

Features of Language

Introduction

Some general conclusions apply to all communication among living organisms. Communication, even in nonhuman animals, has unexpected complexity; it is a form of cooperative behavior; it includes individual recognition and categorization; and it requires the development of associations and thus memory. Furthermore, noise in communication is inevitable. Critical terms, such as information, signal, and noise, require operational definitions (for more on these topics, see Chapter 1).

These conclusions change basic assumptions about the relationship between nonhuman and human communication. Communication by nonhumans is more complex than expected and communication by humans, as presented in the following sections, perhaps less so. There is clearly a large difference in brain size, cognition, and communication between humans and most other animals. Nevertheless, overstating a difference hinders comprehension just as much as understating it.

For centuries people have sought objective criteria to separate humans from other animals, and language has often taken first place among these criteria. Once Darwin and the early ethologists made it clear that nonhuman animals also have elaborate communication, the focus has narrowed to the properties of language that distinguish humans. Each such proposal has spurred students of animal behavior to probe deeper for parallels among nonhuman animals. Some organization in this process came when Charles Hockett (1960) presented a set of 16 “design features,” or distinctive properties, of human languages. Some of these features, such as a vocal-auditory channel (with its concomitants, broadcast transmission, directional reception, and rapid fading), interchangeability, and specialization are easily identified in diverse nonhuman animals. The remaining design features have more problematic parallels in nonhuman animals. Some of the issues they raise invite applications of the general conclusions listed above. The following discussion focuses on illustrative examples, rather than a general review, of these parallels.

Cultural Transmission

Culture is widespread in species with persistent associations of parents and offspring, but it is also prevalent in many species with less complex social behavior. Culture develops when patterns of behavior are acquired by young individuals as a result of experience with older ones. Prevalent examples in nonhuman animals include migration routes, territory boundaries, mating preferences, food selection, and predator recognition. Cultural transmission in humans as well as in other animals includes relatively unconstrained learning. Possibilities for learning have broad scope within wide predispositions. Nevertheless, the study of nonhuman animals has revealed that even impressively open forms of learning have constraints that guide learning in adaptive ways. For rapid acquisition of complex traits within adaptive boundaries, learning within constraints is perhaps the optimal method. Such constraints (or predispositions) can for instance ensure that learning occurs within species, a particularly clear example of which is song learning by oscine birds. In this case, predispositions must affect responsiveness as well as production. Nevertheless, identifying predispositions for acquisition of human language remains contentious.

Animals other than humans rarely, if ever, appear to engage in teaching, in which an experienced individual directs the attention of an inexperienced audience to a task. Cultural transmission instead appears to result predominantly from observational learning by young individuals in the presence of adults performing routine activities (Matsuzawa 1999). A recent experiment with great tits reveals that observational learning can produce persistent cultural traditions in the feeding behavior of birds, even when the tradition is maladaptive (Aplin et al.

2015). In contrast, schools or apprenticeships are perhaps universal in human cultures. Nevertheless, it is possible that much, perhaps most, of human culture is instead transmitted by observational learning.

Cultural traditions in the songs of passerine songbirds, parrots, and hummingbirds illustrate one of the usual consequences of culture, the differentiation of dialects or traditions among nearby populations. All species of songbirds appear to learn at least some features of their songs. With few exceptions these species develop prominent geographical variation in their songs, whereas songs of other species vary only slightly, in line with variation in morphology. When the individuals of a species sing a single song pattern, all those in a limited geographical area often learn the same distinctive pattern. When individuals have repertoires of songs, dialects often intergrade as the frequencies of different acoustic patterns change progressively but incoherently with location (Marler and Slabbekoorn 2004; Kroodsma 2005; Podos and Warren 2007). When individuals sing only one pattern each, the formation of dialects is perhaps simplified. Nevertheless, distinct dialects in the vocalizations of some parrots and cetaceans include coherent repertoires of patterns.

Geographic differentiation of culture depends on the relationship between two periods in an individual's life: when a young individual learns the relevant behavior and when and where it moves before it eventually settles. The formation of distinct dialects requires either that young birds usually settle within the area of their natal dialects or that learning after settling predominates over earlier learning. In the case of coastal populations of white-crowned sparrows in California, individuals usually sing only one pattern, and dialects occupy nearly distinct areas with irregular shapes 2-20 km across. It remains uncertain whether the songs a young sparrow eventually masters are influenced by experience predominantly before or after dispersal. Conversely, it is also unclear whether or not a young sparrow's decision about where to settle is influenced by early experience.

Furthermore, it is also unclear whether or not dialects in birdsong are evolutionary adaptations (Podos and Warren 2007). Adaptation is a result of natural selection, the spread of alleles associated with advantages for individuals' survival or reproduction. In one possible scenario, dialects might result from adaptations for efficient signal detection in different environments. Alternatively, dialects might promote local adaptations for survival or reproduction in general, by restricting gene flow to populations in distinctive environments. Another possibility is that dialects might arise as collateral effects of the evolution of learning, if, for instance, complex learning was in itself important for mate choice. Or dialects might arise as side effects (pleiotropy) of local adaptations for noncommunicative purposes in structures also used for communication, for instance, if the structure of birds' bills adapted to the characteristics of their food but also constrained the kinds of sounds they could produce. In the latter two situations, the formation of dialects would in itself have no influence on individuals' survival or reproduction. In any case, dialects might promote the genetic divergence of populations and ultimately contribute to the origin of separate species.

All of these issues about dialects apply to human cultures as well. The maintenance of human languages and dialects indicates that individuals learn from older individuals within their natal area and then predominantly settle nearby. The increasing frequency of exceptions in recent centuries is presumably changing the geography of human languages. Each individual's choices of mates and places to settle are no doubt to some extent influenced by their natal language. Important but infrequent exceptions are the abduction of individuals and the translocation of populations as a regular consequence of warfare between culturally distinct groups. No doubt the boundaries between languages are influenced by interactions across these boundaries, but differentiation of languages in turn influences the nature of these interactions.

In contrast to birdsong, human language is distinguished by two levels of geographic differentiation. At one level, there are mutually incomprehensible "languages," which in their geographic pattern resemble the distinct dialects of songbirds. At another level, human "dialects" within a language intergrade more or less progressively, like geographic variation in the frequencies of song patterns of birds that have repertoires. Perhaps more study of songbirds would also reveal multiple levels of geographic differentiation.

A related design feature, learnability, refers to the human ability to learn more than one language. In songbirds with vocal dialects, individuals usually respond to dialects other than their own, although sometimes less so to distant dialects (Searcy et al. 1997). Furthermore, just as children are adept at learning any human language, songbirds acquire any dialect of their species-specific song with apparently equal facility. Bilingual individuals also occur among songbirds. Although careful comparisons are lacking, bilingual proficiency is perhaps as frequent as it is in human populations. Bilingual competence is hard to confirm when individuals sing multiple patterns that vary incoherently with location. When dialects are distinct, on the other hand, persistent study often reveals the presence of bilingual individuals. Most white-crowned sparrows that settle near a dialect boundary sing only the pattern appropriate for their dialect, but some individuals near a boundary are indeed bilingual, with two song patterns, one matching each nearby dialect (Baptista 1977). To clarify these parallels between human and nonhuman cultures, the spatial and temporal properties of the transmission of signals and responses, in relation to the dispersal of individuals, need more attention, both in humans and nonhumans.

Despite these open questions about the process, geographical differentiation is one of the salient features of culture, both human and nonhuman. This differentiation requires errors in cultural transmission and thus reveals a crucial effect of noise in communication. Even if cultural traits have adapted to particular social or physical environments and even if migration of individuals introduces novelties in new areas, nevertheless errors must initiate the process of cultural change somewhere. Errors in cultural transmission are analogous to genetic mutations, insofar as both are transmitted to subsequent cohorts, although the rates of innovation and mutation can differ. Migration is similar in the two also, except that migrating individuals can abandon cultural traits but do not change genes (although the expression of genes might change). Cultural and natural selection, on the other hand, lack close analogy. The plasticity of individuals' traits, the rates of transmission, and the recipients all can differ markedly. Nevertheless, in the case of human languages, the stability and restricted acquisition of language might approach that of genetically canalized phenotypes. Close relatives usually play a predominant role in transmitting language. Language proficiency by humans requires years to develop and then, after the lapse of sensitive periods for acquiring some features of language, changes with great difficulty.

The transmission and innovation of culture depend on communication. Unless completely arbitrary, without advantages or disadvantages for signalers or receivers, this communication evolves by the same process of mutual optimization that applies to the evolution of all communication (see Chapter 2). The advantages for signalers depend on the behavior of receivers, and the advantages for receivers depend on the behavior of signalers. Errors, from multiple sources, are inevitable. Even perception evolves by optimization with errors. The crucial role of errors, in other words noise, in all communication and perception means that culture, in humans or other animals, cannot be understood without studying errors. In studying the evolution of birdsong, the variation in learning is as important as the norm. The same is true of all human culture, including language. Yet studies of culture have usually focused on the norms, to the exclusion of the errors. Understanding communication requires attention to exceptions as well as norms.

Semanticity, Displacement, Arbitrariness, and Discreteness

This set of Hockett's design features refers in one way or another to how signals are related to situations. Many animals use discrete signals. Discrete signals are likely to improve discrimination by receivers and thus could have advantages in noisy situations such as long-range communication. In line with this expectation, discrete articulation might be more pronounced in long-range speeches by humans than in close-range conversation. This variation in discreteness would exemplify the scaling of exaggeration in signals with the level of noise, as predicted for the evolution of all communication in noise (see Chapter 2). More study of variation in signals with contexts could clarify this issue.

Arbitrariness refers to separation of a signal from direct resemblance to or evocation by its referent. Warning calls, for instance, do not resemble the sounds of predators. On the other hand, the pitch or tonality of sounds in some cases directly reflects an individual signaler's overall physiological state, such as its tendency to flee, fight, or freeze (Morton 1977; Reby et al. 2005). Arbitrariness is intended to denote a signal's association with a more cognitive internal state, a specific neurophysiological state rather than a general physiological or emotive one. These alternatives, of course, are the ends of a spectrum of possibilities. Each case lies somewhere between a scream of fear and an abstract notion. Human speech conveys information about a speaker's general states, or emotions, in addition to and concurrently with specific states or abstractions. Any signal has some degree of arbitrariness and some degree of abstraction. No doubt human language includes more cognitive complexity than other animals' signals. An unanswered question though is whether or not it might also include more emotive complexity. This continuum between emotive and cognitive applies to the next two design features also.

Note in passing that arbitrary signals in discussions of sexual selection are defined differently. They have zero utility for receivers and, as argued elsewhere, seem highly unlikely in noisy communication (see Chapter 2).

Semanticity and displacement are related to information in signals about the external environment. Some of the best examples of semanticity in nonhumans are warning calls and food calls. Semanticity applies to signals associated with external referents in contrast with those expressing the signaler's internal states. This distinction is just as untenable here as in the previous paragraph. Recall from previous sections that all of an individual's actions are influenced by both its internal and its external state, by both its current constitution and its impinging sensations. Rather than a distinction between signals that are emotive versus cognitive, there is instead a continuum from more emotive to more cognitive. The question is where particular instances of calls lie on this continuum from emotive to cognitive.

Playbacks show that individuals respond appropriately to alarm or food calls even in the absence of an actual predator or food. Furthermore, some birds and mammals have distinct calls for two or more predators that pose different threats, and playbacks of these calls evoke the appropriate responses. The associations with different predators are in some cases learned. Young vervet monkeys reliably produce the correct calls only after prolonged practice (Cheney and Seyfarth 1990). Many birds in contrast easily learn experimental associations of predator calls with various improbable objects. It is thus clear that some nonhuman animals can learn to produce, and to respond to, signals associated with specific external situations, and in some cases these associations are learned easily.

On the other hand, in normal circumstances these calls are produced and evoke responses, only in the presence of actual predators. Displacement, in contrast, refers to separation of a signal and its referent in space and time. This separation requires memory. With displacement, the cognitive rather than emotive associations of signals, and thus their semanticity, often seem clearer. Furthermore, relatively more cognition is suggested by two forms of noise in communication: unreliable and deceptive signals. Adult vervet monkeys, for instance, learn to ignore the unreliable calls of young individuals (or unreliable adults) (Cheney and Seyfarth 1990). Furthermore, some monkeys and birds produce predator warnings deceptively in the absence of a predator, as discussed further below. Errors in communication, the apparent exceptions, provide intimations of cognition.

Displacement is also attributed to the waggle dances of honeybees. These dances in a hive or a swarm indicate the location of food or potential nesting sites by means of two transpositions. The direction to the goal with respect to the sun becomes the direction of a waggle dance on a vertical surface with respect to gravity, and the total energy expended in flight to the goal becomes the instantaneous expenditure of energy in the dance. The levels of arbitrariness and semanticity are both low. The mapping of direction and distance onto the signaler's overall behavior and the limited memory remove this case from typical human cognition. Arbitrariness, semanticity, and displacement again contribute to a continuum between emotive and cognitive behavior. All are widespread in communication. All are disrupted by noise. Yet their use in noise sometimes reveals a degree of cognition.

Prevarication

Hockett's final design features are at the top of his proposals for human specializations. Nevertheless, there are reports of prevarication, or deceptive use of signals, in many birds and mammals. Predator calls, in particular, can serve to distract higher-ranking opponents so that low-ranking ones have a chance to obtain food or matings. Analysis of communication in noise shows that opportunities for such deception are expected in all forms of communication (see Chapter 2).

Nevertheless, linguists often balk at attributing true prevarication to nonhuman animals, in the absence of evidence for the signalers' intention. Intention becomes one of the definientia of deception. Does the signaler expect or plan to deceive the receiver? How does a person expect or plan to take advantage of another person? Presumably the actor anticipates a personal advantage based on anticipation that the recipient is vulnerable. In the case of deceptive communication, the signaler thus anticipates that a particular receiver is likely to respond in a way that yields an advantage to the actor despite a disadvantage to the receiver. This anticipation is tantamount to mind-reading or having a "theory of mind." There is a circularity here: an actor has intentions provided it can read another's mind, and primary evidence that an actor can read a mind is provided by thoughts such as intentions.

Such circularity in discussions of mental phenomenon excludes not only nonhuman animals from a thinking person's mentality but also all other humans. An operational definition of deception, which avoids such circularity, is a system of signals directed specifically (non-randomly) at receivers from which a response, on average, has advantages to the signaler but disadvantages for the receiver. Such signals require some preliminary strategy. Because responding to signals should evolve to increase a receiver's advantage in reproduction or survival, deceptive signals, which have the opposite effect, must in general occur infrequently. Consequently, deception often reveals evidence of a cognitive ability by signalers to adjust the frequency of attempted deceptions by itself and others. For instance, a deceptive signaler might anticipate a particular receiver's probable response; in effect it might read its mind, from the temporal context of signaling.

Reports of deception by nonhuman animals indicate that this behavior is, as expected, usually dependent on the circumstances of the audience. Signals that call attention to the presence or location of food often depend on whether the audience is likely to be a competitor or a partner. Recognition (categorization) of individuals, not just broad classes of individuals, is often critical. Anticipation of the state of the audience is of course widespread in nonhuman signaling, but this attribution applied to specific individuals matches the kind of behavior associated with intentions, as just defined. It presumably occurs in many sorts of social interactions, from anticipatory cooperation to strategic aggression (Cheney and Seyfarth 1990, 2007; Seyfarth and Cheney 2014). It applies to roosters advertising food to attract hens (Gyger and Marler 1988), to subordinate males courting females without drawing the attention of dominant males (Smith et al. 2011), and to territorial warblers challenging specifically those neighbors that have trespassed (Godard 1993). The complexity of an organism's "theory of mind" depends to a large degree on the complexity of its categorization of the individuals it interacts with. Furthermore, intentions in the form of signals adjusted to contexts, whether human or nonhuman, are subject to errors.

Duality of Patterning and Openness

These final, and most problematic, design features are closely related. Openness in human language, an ability to rearrange units of sound to produce new phrases, depends on duality, the organization of speech into at least two hierarchical layers. Elements of a lower layer, which in themselves have little or no meaning, are arranged to produce units of higher layers, which do have meaning. As previously suggested, the operational definition of the "meaning" of a signal, in a particular context, is the receiver's usual response, overt or covert, immediate or remembered. Language is often described with three layers, phonemes, morphemes, and phrases (sometimes with syllables of phonemes interposed), of which only phrases convey meaning. Duality of patterning requires two layers arranged so that recombined discrete components of one layer are nested within components of the other.

Many nonhuman animals have substantial repertoires of discrete signals (see Chapter 2). In those species that produce sequences of signals, some recombine signals into larger performances. Some songbirds use the same components in different sequential patterns of song, but some use different components in each sequential pattern. There is sparse evidence that the different sequential patterns in these repertoires convey different information, for instance, by association with different contexts or internal states, or evoke different kinds of response (Wiley et al. 1994). An absence of evidence is particularly inconclusive here, because finding significant associations with complex patterns of recombining elements becomes statistically challenging. Anthropologists have a big advantage in deciphering a previously unknown human language, as a result of their preconceptions about what humans are likely to talk about. Humans, in other words, have an anthropocentric theory of mind, one that provides much less help in deciphering communication of other species. Humans perhaps should not underestimate other animals.

Despite these possibilities for other species, human language no doubt has remarkable capability to recombine components of signals to convey a vast complexity of information. Nevertheless, components of language are not nearly so distinct in actual speech, in all its various contexts, as they are often presumed to be. Contextual and individual variation in phonemes and morphemes is well known. Furthermore, everyday conversation might depend heavily on phrases as units, learned for production and response as units, rather than as recombined components. All languages are beset with idioms and pat phrases, particularly for routine communication, phrases learned as units rather than by rules. In English, few people know the expressions "thin as a rail" or "what's up," for instance, as anything other than "thin-as-a-rail" or "whats-up," single units of expression, not recombining units. Another example is the instability of prepositions (and grammatical cases). Across and within languages, association of prepositions with contexts often defy consistent definition and instead become erratic or idiosyncratic. Most of these associations are presumably learned and deployed as units without parsing. Proficiency thus might often depend on mastering associations of these unitary phrases. Even when some parsing of recombined components is necessary, the associations of phrases can depend on common underlying metaphors (Lakoff and Johnson 1980). Indeed, morphemes raise the same questions. Many words are sequences of two or more syllables, which in combination evoke unitary associations. Etymological stems for syllables are rarely parsed. Even then underlying metaphoric associations dominate. Thorough study of the variability of human speech at all levels might reveal that duality of patterning requires a dose of grammatical fantasy.

The deployment of writing in itself changes communication. Writing allows much greater permanence than does neural memory alone, and this permanence allows a reader, as a receiver of signals, to examine and even to review the structures of phrases more carefully than is possible in conversation. With time available, humans can indulge their drive to categorize, by abstracting, cataloguing, and eventually prescribing patterns of usage. In reality both human and nonhuman communication are permeated with unexpected variation, as a result of errors in production, transmission, or reception of signals, but also just idiosyncrasies in usage by individuals or small groups of communicating individuals. The irregularities, idioms, idiosyncrasies, and errors are an inescapable part of language.

Openness is an abstraction or exaggeration of reality as well. Although linguists often claim infinite possibilities for language, the components of language are finite, the human brain is finite, and the practical possibilities for combinations are finite. Speaking humans do not produce stereotyped phonemes that recombine to form stereotyped morphemes and then phrases with unlimited possible meanings. The number of possible associations is no doubt large, but the number of associations humans make in using language might not exceed the number of associations they make in categorizing other humans (Wiley 2013). Errors also limit associations of attributes with other objects.

Furthermore, limited evidence is a temptation for simplification of nonhuman animals' behavior. For instance, songbirds are usually thought to have repertoires of distinct patterns of notes, from one to several hundred such patterns. Careful inspection, however, reveals much variation in details, little or none of which has any current explanation. Perhaps even greater complexity in sequences of sounds occurs in cetaceans.

The difficulties of investigating complex signals include the impediment of lacking an appropriate theory of mind on which to base hypotheses. Nevertheless, much variation is no doubt meaningless, just as it presumably is in human speech.

A promising way to investigate variation consists of looking for dependence in recombinations of components. A simple example is provided by displays of the Carib Grackle *Quiscalus lugubris* (Wiley 1975). Males perform conspicuous displays to females and other males. Each display involves raising wings, tail, and bill to varying degrees. Elevated wings show some association with displays toward females, elevated bills with those toward males. Because these displays are easily observed, a large number can be scored for each element. Analysis reveals that wing and tail elevation are independent of each other, so in this case this species has the potential to generate infinite gradations of wing and bill elevation, each presumably related to neural and mental states while interacting with females and other males. This example of insipient duality and openness indicates how difficult it can be to decide whether complex variation in signals is relevant, erroneous, or simply adventitious.

Hierarchical Organization

As Awareness of the complexities of nonhuman communication has accumulated, attention has focused on hierarchical organization as the key to the relative openness of language. Although responses to “sign stimuli,” simple sensations, occur widely, especially in the initial responses of young organisms or in the quick avoidance of predators, nevertheless, responses to more complex stimulation are also widespread. For instance, many vertebrates, but not humans, have neurons that act as movement detectors, even at low levels of sensory processing. Recognizing patterns is not an unusual capability of many other animals. Although birds and mammals soon after hatching or birth have reflex (highly canalized) responses to simple stimulation associated with predators, they often quickly learn more complex associations. Object constancy, an ability to recognize (form associations with) a set of sensations as a unique object despite varying perspective and occultation, does not differ in principle from other forms of recognition, including recognition of individual conspecifics or recognition of verbal sequences. Recognition of patterns, spatial or temporal configurations of components, is thus a mental capability that occurs widely in animals as well as humans.

Much recognition is potentially hierarchical. Any particular instance of a set of sensations could be recognized as belonging to one or more progressively more inclusive and complexly embedded or overlapping categories. A territorial neighbor can be recognized, for instance, despite singing multiple different song patterns, at different locations and distances, under different environmental conditions. Furthermore, it might be recognized more specifically as one that had recently trespassed or one known from a previous year (Godard 1991, 1993; Wiley et al. 1994; Godard and Wiley 1995). Hierarchical categories are recognized by Aristotelian definition, with consistently defined features. Alternatively, categories might be recognized by family resemblance, with inconsistently shared features. In the first case, all members of a category would share an inclusive set of features, as in a phylogenetic tree. Membership in categories would be unambiguous. In the second, members of a category would each share some but not necessarily the same set of features with every other, as in an actual family. Ambiguity might occur.

Chomsky (2005) recognized the importance of categorization when he proposed that merging is the crucial cognitive operation of language. Merging, in the usual sense of simple combining, is nevertheless too simple for his examples, which require combining elements from two separate categories, subjects and predicates, in order to produce a phrase with meaning. In more physiological terms, Chomsky's merging is not just association of perceptions but association of elements from two categories of perception. Associating perceptions from two categories is cognitively similar to associating individuals with different contexts. It is a matter of common experience that contexts affect recognition of individuals, presumably because objects such as individuals become associated with their contexts. Thus each context merges, in Chomsky's sense, more easily with some individuals than with others, just as each verb merges more easily with some nouns than others. In this way, parsing social interactions might require cognitive abilities that could be coopted for language. Yet it is not clear whether the relevant categories are recognized by definition or by family resemblance. Furthermore, both in language and social interaction, associations might sometimes be recognized as units, without any parsing, in other words, without any analysis and merging of parts, at all.

Consequently it seems unlikely that either language or social interaction is organized entirely hierarchically. Nevertheless, this particular form of organization has received special attention as a possibly fundamental feature of language. Hierarchy connotes two distinct kinds of organization. Human institutions are hierarchical in the sense of a chain of command, with each individual overseeing a set of subordinates. The organization is like a multidimensional pyramid, with lower sets embedded (nested) within higher sets. In contrast, tennis players and perhaps society mavens are ranked unidimensionally, in a ladder. In nonhuman social organization, dominance hierarchies take the latter form. An example of such a pattern is (A>B>C>D) where A through D are individuals ranked on a single dimension. Embedding in this case consists of pairs of closely ranked individuals inserted between pairs of distantly related individuals, (A>(B>C)>D). Human language is also linearly ordered in time (speech) or space (writing), with similar nestings of components, where A through D are words (morphemes). Several experiments have suggested that birds can recognize sequences like (AABB) or (ABBA), although it is not clear that they can generalize such a pattern to new exemplars (Van Heijningen et al. 2009). The second sequence superficially matches a pattern of embedded phrases in language. On the other hand, the first, not the second, sequence matches complex dominance hierarchies that result when individuals' rankings are embedded within families' rankings or when dominants create coattails for familiar subordinates. There are indications that baboons and birds can "parse" a dominance hierarchy with embedded clusters (see Chapter 2).

Embedding of phrases in language (often confusingly termed recursion) is more complicated than either diagram above. Each phrase consists of components with different roles, for instance, nouns and verbs, or more generally objects and attributes. The exact relationship of these two types of components in a phrase is marked either by their sequence or by tags (inflections), and the relationships of phrases are also marked by their sequence or by tags (conjunctions). A more accurate diagram of embedding in language is thus (A1(B1B2)A2), where A and B are phrases and 1 and 2 designate appropriate objects and attributes within each phrase (Corballis 2007). A human receiver associates A1 with A2 and B1 with B2. These associations are either temporal (for a listener) or spatial (for a reader). The cognitive issue is whether or not nonhuman organisms can respond reliably to (A1–A2) regardless of whether these two components are separated in time or space by analogous phrases, such as (B1B2).

Sensations evoking responses by animals are often (perhaps always) composed of multiple elements in particular arrangements. Thus it is not surprising to find that animals can master associations like (A1–A2) regardless of some interruptions. An ability to respond to the associated sensations despite interruption comes close to object constancy, discussed above. The experiments mentioned above, which show that birds fail to generalize such patterns (Van Heijningen et al. 2009), perhaps miss the point. Object constancy is probably not generalized either; instead each object is learned by family resemblance of its particular features, despite various interruptions, and eventually evokes a unitary response. Furthermore, as discussed above, it is not clear that using language requires parsing of components.

Embedding in language is even more complex, because in this simple case, (B1B2) modifies A1, so that (A1B1B2A2) becomes in effect ((A1(B1B2))A2). The phrase “The cat that the dog attacks hisses” does not merely merge two phrases “The cat hisses” plus “The dog attacks.” Instead the inner phrase changes the meaning of the outer phrase; the conjunction makes the connection. Can some nonhuman animals recognize an association of two signals, each of which associates components from two categories, regardless of whether one signal is interposed between the components of the other signal? Perhaps.

As important as embedding is for human language, the relationship established by an action also seems critical, as in a phrase such as “The dog attacks the cat.” The relationship between the “dog” and “cat” is in part specified by the action “attacks” but also by marks that indicate the relationship of each object to the action. This latter relationship is marked in English primarily by sequence: (A1 B1 A2). In Russian, and many other languages, the relationships are marked by modification of each noun (the case of each noun, either nominative or accusative): (A1n B1 A2a). In Russian, sequence has less salience (and determinatives are usually absent), so this inflected phrase would elicit a similar response in the sequence (A2a A1n B1) or in any other sequence. Can animals recognize a three-component signal in which the components have particular relationships specified either by arrangement in time or space or by at least one modifying (case) component. Despite the complexity, this challenge is nevertheless met by some nonhuman animals (Herman and Richards 1984; Marino et al. 2007).

Language as Criteria for Responses

The forgoing discussion has failed to identify a key to language. It has made little progress in isolating any qualitative requirement for the use of language that nonhuman animals do not already have, in some cases to a considerable degree. Yet it seems clear that no other species engages in communication approaching the complexity of human language. Nor have they achieved the levels of technological competence that language has catalyzed among humans. Before suggesting a solution to this paradox, this section first provides a different way to conceptualize the use of language. Recent thinking about language has usually started from the top, from idealizations by grammarians and linguists (Hauser et al. 2002; Chomsky 2005; Tomasello 2010; Fitch 2017; Seyfarth and Cheney 2017). The following starts from the bottom, from basic neural mechanisms for all communication.

Language consists of clusters of perceptions. Categorizing clusters of perceptions to form components of language is fundamentally the same as categorizing sensations to form primary perceptions. Sensations have inherent variability, as a result of variation introduced by their sources (including human signalers, speakers, writers, or signers), their receivers (including human listeners, readers, or sign readers), and the medium in between. This variability is noise in the perception of signals. All receivers of signals in noise make decisions to associate sensations with responses by means of criteria for response (Wiley 2015, 2017; see Chapter 2). These criteria associate particular sets of incident sensations with particular sets or levels of responses, either overt or covert, in action or memory. Classification of sensations is thus a result of their associations with responses. Initial perceptions are the first responses to sensations.

The criteria for each decision, like all other features of living organisms, develop in the course of each individual's life as a result of an interaction between its genetic constitution and its environmental conditions. At any moment these criteria depend on the individual's current physiological and anatomical state. The association of sensory input with response thus results from the individual's current state and the impinging sensations.

Classification of sensations is the preliminary stage in the eventual classification of perceptions into the components of language. Initial perceptions are the fundamentally meaningful categories of sensations. At each subsequent stage of categorization, the process of association divides perceptions into progressively more specific perceptions or other responses. At each stage, the criteria for categories might combine definitions and family resemblances, and the criteria might change, with developing familiarity, from sequential or inflectional parsing of relationships among components to immediate unitary detection. The inherent variation in sensations at the root of the process propagates into variation in perceptions at every higher stage. Noise permeates all stages in the processing of language. It requires neural decisions to recognize categories of sensations or perceptions at each stage. The variation and exceptions are as important as the norms.

This perspective of language does not preclude human cognitive criteria that quantitatively exceed those of nonhumans. Yet it has not identified a qualitative cognitive capability that nonhumans entirely lack. Combinations of associations in time and space admit great complexity. Increased complexity no doubt can lead to a "great leap forward." It is a truism that quantitative change can lead to qualitative change. Indeed humans have a nearly inexhaustible impulse to categorize, so any change, no matter how small, can invite categorization as a qualitative change. Yet it is not clear whether or not such a "leap" requires any innovations beyond one small step at a time on the same path. Nevertheless, something extraordinary happened when human language developed. It might not have required any advance in cognition.

From Nonhuman to Human Language

Chimpanzees, bonobos, bottlenose dolphins, grey parrots, and other nonhuman animals have demonstrated surprising capabilities (SavageRumbaugh and Lewin 1994; Marino et al. 2007; Pepperberg 2004). They easily master the use of abstract symbols, reference, displacement, and sequential recombinations. They can both produce and respond appropriately to symbols. They can use symbols to communicate with conspecifics to solve problems. They can to some extent acquire these capabilities from other conspecifics.

Despite some animals' language-like abilities, there is no clear evidence that these abilities are used for communication in natural situations. It is interaction with language-capable humans that reveals the inchoate capabilities for language in apes, parrots, and dolphins. This situation recalls a repeated result of all mathematical models for the evolution of communication. The initiation of communication, of whatever simplicity or complexity, must surmount a hurdle. Rare alleles associated with response to a new signal are unlikely to spread in the absence of individuals producing the signal. Conversely, rare alleles associated with producing a new signal are unlikely to spread in the absence of individuals responding to the signal. As seen in the models for the evolution of communication in noise and for the evolution of traits and preferences by sexual selection, the benefits of signaling depend on responses, and the benefits of responding depend on signals. The frequencies of signals and of responses must reach some threshold before the benefits of signaling and responding begin to spread (see Chapter 2).

This conclusion invites application to these language-capable but language-deficient populations. Chimpanzees have evolved enough mental capabilities to provide the advantages of language, yet in natural situations language among chimpanzees is, by all evidence, absent. There could well be advantages for apes to have language to assist in coordinating cooperation within their groups and competition between groups. They appear to have enough of a start in mental competence. Perhaps, so far, neither the frequency of gestures nor responses have reached the necessary threshold. Who knows? Perhaps a fortunate coincidence, just one small group with by chance a few gesturers and a few responders, a few individuals with just the requisite predispositions, such as Kanzi seemed to have, and overnight a new kind of chimpanzee society based on new possibilities for communication would sprout. Subsequent natural selection would enhance these incipient predispositions for specialized learning.

After that leap forward, natural selection would result in evolution toward optimal signaling and responding. Perhaps it would evolve rapidly, accelerating as expected for sexual selection but also for any frequency-dependent selection of mutualism in communication. The increase in size and complexity of society with the introduction of agriculture might contribute to selection for greater complexity in language. The concurrent invention of writing would almost certainly increase the potential complexity, as a result of the greater storage and review of language. Then there would come printing and eventually the Internet and computers to assist with storage, search, translation, and associations of language. Perhaps even tools and fire might have affected the evolution of language, or, perhaps more likely, language affected them.

It is easy to imagine how increasing competence with language could improve thinking, which is after all internal communication. It presumably would require higher levels of association and memory. In contrast, a proposal that language-like thought might precede language for external communication with other individuals seems unlikely. Private language would lack the stability acquired from consilience in the process of communication with other individuals (Wiley 2015). Instead the evolution of language is likely to have promoted the evolution of thinking.

Conclusion

This scenario supposes that the advantages of complex hierarchical societies with some incipient forms of cooperation and monitoring of other individual's social relationships might have favored the initial evolution of advanced mental capabilities. The advantages of multiplicity and specificity in individual recognition might be enough to promote the evolution of complex associational learning. The requirements for criteria based on complicated family resemblances and for object constancy in challenging conditions might produce enough cognitive complexity. The ultimate form of cooperation, language, would then just need the impetus to get past the impasse of signalers without receivers and receivers without signalers. A boost in frequencies of signals and responses might come with a random perturbation during a bottleneck in population size. After crossing the initial hurdle, natural selection on the predispositions for language could take hold. Perhaps faster than so far imagined, the use of language would flourish.

References

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, 518, 538–541.

Baptista, L. F. (1977). Geographic variation in song and dialects of the Puget Sound white-crowned sparrow. *Condor*, 79, 356–370.

Cheney, D. L., & Seyfarth, R. M. (1990). *How Monkeys See The World: Inside the Mind of Another Species*. Chicago: University of Chicago Press.

Cheney, D. L., & Seyfarth, R. M. (2007). *Baboon Metaphysics: The Evolution of a Social Mind*. Chicago: University of Chicago Press.

Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36, 1–22.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science*, 31, 697–704.

Fitch, W. T. (2017). Empirical approaches to the study of language evolution. *Psychonomic Bulletin & Review*, 24, 3–33.

Godard, R. (1991). Long-term memory of individual neighbours in a migratory songbird. *Nature*, 350, 228.

Godard, R. (1993). Tit for tat among neighboring hooded warblers. *Behavioral Ecology and Sociobiology*, 33, 45–50.

Godard, R., & Wiley, R. H. (1995). Individual recognition of song repertoires in two wood warblers. *Behavioral Ecology and Sociobiology*, 37, 119–123.

Gyger, M., & Marler, P. (1988). Food calling in the domestic fowl, *Gallus gallus*: The role of external referents and deception. *Animal Behaviour*, 36, 358–365.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Herman, L. M., & Richards, D. G. (1984). Comprehension of sentences by bottlenosed dolphins. Cognition, 16, 129–219.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal Sounds and Animal Communication* (pp. 392–430). Washington, DC: American Institute of Biological Sciences.

Kroodsma, D. (2005). *The Singing Life of Birds*. Boston: Houghton Mifflin Co.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago:

University of Chicago Press.

Marino, L., Connor, R. C., Fordyce, R. E., Herman, L. M., Hof, P. R., Lefebvre, L., ..., & Rendell, L. (2007). Cetaceans have complex brains for complex cognition. *PLoS Biology*, 5, e139.

Marler, P., & Slabbekoorn, H. (2004). *Nature's Music: The Science of Birdsong*. Amsterdam: Elsevier Academic Press.

Matsuzawa, T. (1999). Communication and tool use in chimpanzees: Cultural and social contexts. In M. D. Hauser & M. Konishi (Eds.), *The Design of Animal Communication* (pp. 645–671). Cambridge, MA: MIT Press.

Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, 111, 855–869.

Pepperberg, I. M. (2004). Grey parrots: Learning and using speech. In P. Marler & H. Slabbekoorn (Eds.), *Nature's Music: The Science of Birdsong* (pp. 364–373). Amsterdam: Elsevier Academic Press.

Podos, J., & Warren, P. S. (2007). The evolution of geographic variation in birdsong. *Advances in the Study of Behavior*, 37, 403–458.

Reby, D., McComb, K., Cargnelutti, B., Darwin, C., Fitch, W. T., & Clutton-Brock, T. (2005). Red deer stags use formants as assessment cues during intrasexual agonistic interactions. *Proceedings of the Royal Society of London B: Biological Sciences*, 272, 941–947.

Savage-Rumbaugh, S., & Lewin, R. (1994). *Kanzi: The Ape at The Brink of The Human Mind*. New York: John Wiley and Sons.

Searcy, W. A., Nowicki, S., & Hughes, M. (1997). The response of male and female song sparrows to geographic variation in song. *Condor*, 99, 651–657.

Seyfarth, R. M., & Cheney, D. L. (2014). The evolution of language from social cognition. *Current Opinion in Neurobiology*, 28, 5–9.

Seyfarth, R. M., & Cheney, D. L. (2017). Precursors to language: Social cognition and pragmatic inference in primates. *Psychonomic Bulletin & Review*, 24, 79–84.

Smith, C. L., Taylor, A., & Evans, C. S. (2011). Tactical multimodal signaling in birds: Facultative variation in signal modality reveals sensitivity to social costs. *Animal Behaviour*, 82, 521–527.

Tomasello, M. (2010). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Van Heijningen, C. A., De Visser, J., Zuidema, W., & Ten Cate, C. (2009). Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National*

Academy of Sciences, 106, 20538–20543.

Wiley, R. H. (1975). Multidimensional variation in an avian display: Implications for social communication. *Science*, 190, 482–483.

Wiley, R. H. (2013). Specificity and multiplicity in the recognition of individuals: Implications for the evolution of social behavior. *Biological Reviews*, 88, 179–195.

Wiley, R. H. (2015). *Noise Matters: The Evolution of Communication*. Cambridge, MA: Harvard University Press.

Wiley, R. H. (2017). How noise determines the evolution of communication. *Animal Behaviour*, 124, 307–313.

Wiley, R. H., Godard, R., & Thompson, A. D. (1994). Use of two singing modes by hooded warblers as adaptations for signaling. *Behaviour*, 129, 243–278.

Chapter 3

Evolution and Self-awareness

Introduction

It seems clear to me that my subjective mental states are my own. Furthermore, this act of introspection shows that I have some ability to think about my mental states. I would say that I am self-aware. Furthermore, my experience indicates that most humans have such mental states of their own, including self-awareness. The question thus arises where do these subjective mental states, including self-awareness, come from? What causes or explains their presence and content? Has this human capability evolved?

From the earliest times, discussions of consciousness have been closely related to those of free will. Volition is still for many people a definitive attribute of consciousness, which in turn is a definitive attribute of humans. Both volition and consciousness, in turn, are closely allied with rational thought. Often these three capabilities are combined as the distinctive characteristics of the human soul. For many people, the attributes of these capabilities seem entirely apparent by introspection. For all of these reasons, discussions of self-awareness and volition have often become doctrinaire, for instance, in both Judeo-Christian-Islamic-Vedic morality and Brahmanic-Buddhist-Taoist-Gnostic transcendence. Human and nonhuman animals have seemed to occupy their respective rungs on a *scala naturae*, each of which had its distinctive attributes augmenting the rung below and subsumed by the rung above. Consciousness and its concomitant volition and rationality were the attributes that separated humans from all lower forms of life.

Modern western philosophical and theological discussions began in the sixteenth and seventeenth centuries with fundamental controversies between Christian sects about divine grace and predestination and with Descartes' subsequent dictum "*Je pense donc je suis*." In recent centuries, contrasting positions have developed between environmentalist and nativist theories of human perception (or between sensory and mental determinants of thought). These trends have culminated in an emphasis either on culture or on intrinsic structure as the predominant determinants of consciousness and language. In all of this history, because it was assumed that the uniqueness of humans depended on consciousness and self-awareness, there was little attention to the possibility of their evolution.

Continuity between Human and Nonhuman Animals

Darwin's revolutionary book, The Expression of the Emotions in Man and Animals (1872), for the first time, proposed some continuity between the minds of humans and other animals as indicated by their respective behaviors. Since then, studies of learning and instinct, as well as the ontogeny and phylogeny, of behavior have had progressively increasing influence on thinking about consciousness and awareness. Particularly relevant have been experiments that explore the limits of nonhuman intelligence. These experiments have raised difficult questions about the relationships among language and thought and learning and consciousness.

For instance, studies of a number of nonhuman animals (especially chimpanzees, other apes, parrots, and dolphins) have indicated that these organisms can respond to complex stimulation, such as encoded queries and requests, in ways that resemble our own use of language (Savage Rumbaugh et al. 1998). Just as we routinely attribute consciousness and will to other people by empathy based on their behavior, we are inclined to attribute these capabilities to the subjects of these experiments. These studies are unusual among biological and psychological experiments in two ways: their small samples of subjects and their intensive involvement of humans (as opposed to experiments with many comparable subjects and largely mechanical interactions with them). The small sample is justified because the interest is in the potential rather than the norm of behavior. If one individual chimpanzee can converse like a human, that alone makes the point.

The lack of replication in combination with intensive human interactions with subjects raises further issues. Each of these studies has been conducted in one laboratory and directed by one principal investigator or a small team. No team has ever carefully replicated the procedures of any other. Nevertheless, there has been an accumulation of similar results with similar protocols, so that it is becoming more difficult to exercise broad skepticism about the kinds of responses evoked.

There is an even deeper controversy about language-like responses by nonhuman animals. Interpretations of these responses often devolve into a polarity between attributing the observed responses to no more than thorough (rote) learning as opposed to the spontaneity or creativity that human language seems to show (Lutz 2009). There are twin problems here – it is difficult to distinguish complicated from random patterns of responses, and it is also hard to distinguish repeated from rote responses. Infrequent, unprecedented, and unrepeated responses, just what we look for in consciousness, cannot be easily shown statistically to differ from accidental or random responses. On the other hand, statistically significant patterns of response require replicated results, which then come to resemble thoroughly learned or rote responses.

The issue of thorough learning, as opposed to volitional thought, is a pervasive problem in comparative studies of consciousness. Consider another example of convergence in mental abilities of human and nonhuman animals – abilities to respond to oneself in a mirror and to attribute mental states to others (and by extension therefore to oneself). Responses to mirrors are no doubt tricky. Many animals respond to mirror images (itself a remarkable capability) as if the image were another individual of its species, perhaps a rival evoking aggression. Chimpanzees in contrast behave as if they see themselves in a mirror, for instance, by touching unexpected marks on their faces that they see only in a mirror. Such responses to mirrors, which seem normal to most humans nowadays, indicate a remarkable advance in mentality. Yet this ability clearly requires learning. Humans with no experience of mirrors or photographs of themselves, as anthropologists often report, do not easily recognize their images. And anyone can try the experiment of directing movements (or even identifying oneself) in a mirror image of a mirror image of yourself, an experience that reveals uncomfortably that mirrors require considerable practice to master. Even recent experiments in which chimpanzees without a direct line of sight can use a mirror to sign to a human recipient, but do so only when the recipient is looking at the mirror (Lurz et al. 2018), are subject to the same questions about thorough learning. Evidently humans nowadays master some superordinate associations involving mirrors that most nonhumans have not, but humans have not mastered all the possible associations. Nor is it easy to determine whether or not some nonhuman animals have comparable mastery (De Veer and van den Bos 1999).

Actions that result from rote learning raise another question about consciousness. Humans often master well-practiced tasks to the extent that they are performed unconsciously. Conspicuous examples are actions or sequences of actions performed routinely, such as making a cup of coffee each morning. Yet on any one occasion, the awareness of having performed an action can escape us, for instance, when we are perplexed about whether we had already added the sugar. Certain actions, swallowing and walking, for instance, consist of complex muscular coordinations that we are seldom aware of, although this lack of awareness can cause serious accidents.

Furthermore, this issue of the criteria of consciousness merges with the issue of private sensations. How can we know, other than by imputations based on coarse empathy, what another organism feels or even senses? Introspection is the source of these insights. How can it be determined that another organism, other than myself, is conscious, acting by will rather than rote, in the same way I do?

Neurophysiology of Consciousness

An obvious possibility for recognizing consciousness is to investigate neural activity during presentations of stimulation. Neurobiologists can now detect in detail the neural events that result from sensory stimulation, both those sorts that we normally are aware of, for instance most exteroceptive sensation, and those we are normally not aware of, for instance most proprioceptive sensation. This approach can extend to neural events in the brain. Physiologists now know a great deal about local areas in the brain, even particular neurons, that are specialized for analyzing sensations, controlling muscles, generating emotions, consolidating memories, comprehending or producing language, even recognizing a visual pattern as a human face. Centers in the brains of nonhuman animals have similar functions (the centers for learning and producing songs by birds are particularly well documented). Is there such a center for consciousness?

Suggestive in this case is an experiment that seems to reveal a half-second or so delay between the initiation of a spontaneous action, on one hand, and awareness of it, on the other (Libet et al. 1983). Because action precedes awareness, it appears that action triggers awareness, rather than vice versa, so that consciousness is the effect of our actions rather than their cause. Volition becomes an illusion, and consciousness seems to be a distinct operation, a candidate for localization in the brain. Some qualifications are in order, though. Subjects record their awareness of the action by remembering the exact position of a spot moving rapidly around the face of a clock. Recording this visual stimulus in memory is itself a response to the spontaneous decision to act, just as much the act itself is a response to this decision. Such an experiment thus does not necessarily reveal that action precedes volition. Instead each of these two operations requires different neural events lasting finite, and evidently not exactly equal, amounts of time. Furthermore, the memory is encoded in language, which becomes the sole means of obtaining the datum actually recorded by the experimenter. Again we are back to questioning how we can know what another organism feels or thinks, unless that organism tells us in some way.

The inability so far to find a locus in the human brain specialized for consciousness has led to proposals that awareness results from distributed networks of neural interactions. Computer programs for learning complicated tasks often employ "neural networks," one or more intermediate layers of "cells" that reciprocally influence each other's activity, between an input (sensory) layer that provides initial conditions and an output (motor) layer that represents the response. Further programming defines the utility of any response and uses this evaluation to regulate the stability or variability in the properties of cells in the intermediate layers. An arrangement that produces responses with low utility is thus allowed to change (mutate) before subsequent trials; those that lead to responses of high utility are stabilized (saved), in other words, learned. Although these computer programs are called "neural networks," it is still not clear how closely they resemble operations in a brain. Only at a superficially general level can we suppose that distributed operations in the brain share the features of computational "neural networks." The specifics of neural processing to produce consciousness remain as elusive as ever.

Continuity between Brains and Other Machines

The relationship between consciousness and language arises in proposals to distinguish humans from other machines – or by extension to determine whether or not any machine is conscious. Turing's test and Searle's modification of it are examples (Dennett 1991; Searle 1997). Each involves a judge posing problems to unseen contestants. The issue is whether a human (conscious) contestant can be distinguished reliably from a nonhuman (unconscious) one. Searle contends that Turing's test would not distinguish between a human who understood a language and one who just followed rules by rote. It thus could not distinguish a conscious human from an unconscious machine. By extension, it is worth emphasizing, it would also not distinguish between a conscious and an unconscious machine. A fundamental question here is whether or not conscious behavior, such as language, is strictly rule-following or not. And thus whether or not humans are strictly rule-following machines or not. We might also extend this question to whether or not machines do or do not strictly follow rules.

Noise as a Determinant of Consciousness

These conundrums about the relationships of volition, consciousness, and learned and unlearned behavior all intertwine with issues of language and even communication in general. Any evidence about an organism's consciousness depends largely, in the final analysis perhaps exclusively, on what it reveals in its behavior. The evolution of consciousness thus depends in a fundamental way on the evolution of communication. It is thus remarkable that noise influences the evolution of communication in a way that provides a straightforward explanation for the evolution of subjective experience (Wiley 2015).

Noise, as Claude Shannon first emphasized in his revolutionary analysis of information, is anything that results in errors by receivers during communication (Shannon and Weaver 1963). Noise can consist of extraneous irrelevant background stimulation that mixes with signals during transmission from signaler to receiver. This is what is commonly thought of as noise. Noise, in Shannon's sense of errors by receivers, can also result from degradation and attenuation of signals during transmission. It can also result from irregularities in a signaler's nervous system which introduce irregularities in its signals; and it can consist of analogous irregularities in a receiver's nervous system which introduce irregularities in its perceptions. As a result of any of these sources of noise, a receiver's perception only imperfectly reflects a signaler's actual situation.

To apply this approach to the evolution of consciousness, note that the dilemma confronting a receiver of signals in communication is strictly analogous to that confronting a perceiver of external objects and events in general. Noise in perception can result from mixing of sensations from irrelevant sources, from degradation and attenuation of stimulation during transmission, and from unpredictability in a perceiver's own nervous system.

At the moment of perception, a perceiver has no way to determine whether or not the perception corresponds to a particular external situation or to an erroneous illusion. All the perceiver knows at the moment is its perception. Nevertheless, memory of repeated perceptions, especially in combination with communication with other individuals, could reveal these discrepancies. In this way such an organism, capable of thought and language, could develop a sense that its own perceptions differed, in some respects and on some occasions, from those of others. Both some ability for abstract thought (a capability for generalization and discrimination) and some ability for communication of such abstractions seem crucial for this awareness. Otherwise individuals would be isolated within the shell of their own perceptions. They might well learn to avoid or to prefer certain perceptions, but it would be difficult to compare them with other individuals'.

Because of noise in perception or communication, a perceiver or receiver must make a decision every time it acts on any sensation. It must decide whether the sensation warrants a response (and also which response). In other words, it must decide whether a sensation is a signal (with some relevance for the perceiver) or noise (with no, or misleading, relevance). Noise creates the unavoidable possibility of two incompatible kinds of errors in perception, false alarm or missed detection. All perceivers, even those organisms such as sponges or bacteria with no nervous system like ours, are perforce decision-makers.

Signal detection theory (Macmillan and Creelman 1991), based on Shannon's theory of information, and decision theory allow a formal mathematical analysis of the performance of any perceiver in the presence of noise. Because of the two conflicting sources of error, a perceiver is in a double bind. It cannot reduce one source of error without increasing the other. As a consequence, it can only optimize its decision in particular circumstances and cannot attain perfect performance. This optimization leads to a fundamental conclusion that perceivers cannot escape from noise; their only option is to optimize their responses in each situation.

This is what organisms actually do. All must deal with unpredictable contingencies. Evolution by natural selection provides a mechanism that can optimize, within limits, neural capabilities to make decisions that promote survival and reproduction for the organism. In a fundamental way, nervous systems are decision-making organs devoted to this task of responding efficiently to conflicting possibilities of stimulation. Every organism must confront its subjectivity with some decisions, no matter how crude the mechanism.

Awareness of subjectivity in perception, however, requires a nervous system to form higher-order associations. When these connections between subjectivity and objectivity reach awareness, we can expect consciousness. The process requires a sufficiently complex nervous system. The logical inconsistency of self-reference might indicate that no such system can ever be completely aware of all its operations, even its degree of self-awareness.

Humans clearly have achieved the highest performance so far. Yet the evolution of this capability seems likely to have emerged gradually by successively more complex mental associations. Whether or not other organisms (great apes and bottle-nosed dolphins come to mind as possibilities) have reached states of consciousness comparable (although perhaps not identical) to those of humans and whether or not some future organism or some other deterministic machine might eventually reach higher levels of self-awareness are questions for the future. Perhaps humans have reached an adaptive peak in the evolution of consciousness, so further advances might require brains evolved in a new anatomical/physiological direction.

Conclusion

The mathematical analysis of optimal behavior in noisy situations thus indicates that (1) noise is an inescapable component of communication, (2) subjective awareness of self is a higher-order association of perceptions and responses, (3) decision-making is a fundamental component of all communication and perception, and (4) both processes are as unpredictable as the unavoidable noise. An advantage of this analysis of the evolution of communication in noise is the framework it provides for addressing the questions posed at the start of this essay.

To account for the source and the content of self-awareness, previous discussion has always relied either on supernatural intervention or on vague neural operations on purified sensations. Supernatural intervention of course obviates any mechanistic explanation, including evolution. Response to pure sensations, on the other hand, leaves each organism encased in its own perceptions, without a way to distinguish between subjective and objective events. The evolution of noisy communication, in contrast, shows that self-awareness (consciousness) results directly from the operations of nervous systems exposed to noisy sensations. We can expect that the neural correlates of self-awareness will depend not only on sensations of interest to an organism but also on the noise mixed with them. The resulting explanation for self-awareness requires no unnatural or unspecified components.

The principal conclusion from such an analysis is that the problems of consciousness might reduce to problems of evolution, signal detection, and neurobiology, all highly mathematical and physical, and thus mature scientific fields. Discussion of the mechanisms of consciousness might therefore migrate from philosophical to scientific discourse.

References

De Veer, M. W., & van den Bos, R. (1999). A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, 58, 459–468.

Dennett, D. C. (1991). *Consciousness Explained*. Boston: Little, Brown and Company.

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activities (readiness potential); the unconscious initiation of a freely voluntary act. *Brain*, 106, 623–642.

Lurz, R., Krachun, C., Mahovetz, L., Wilson, M. J. G., & Hopkins, W. (2018). Chimpanzees gesture to humans in mirrors: Using reflection to dissociate seeing from line of gaze. *Animal Behaviour*, 135, 239–249.

Lutz, R. W. (Ed.). (2009). *The Philosophy of Animal Minds*. Cambridge: Cambridge University Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide.* Cambridge: Cambridge University Press. Reprint, Mahwah: Erlbaum, 2004.

Savage-Rumbaugh, S., Shanker, S. G., & Taylor, T. J. (1998). *Apes, Language, and The Human Mind*. Oxford: Oxford University Press.

Searle, J. R. (1997). *The Mystery of Consciousness*. New York: New York Review of Books.

Shannon, C. E., & Weaver, W. (1963). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Wiley, R. H. (2015). *Noise Matters: The Evolution of Communication*. Cambridge: Harvard University Press.

Chapter 4

Evolution of Free Will (Decision)

Introduction

It has always been accepted that free will, an ability to choose among alternative actions or beliefs, is characteristically human. It seems to result from rational thought or perhaps is simply one aspect of rational thought. Because no nonhuman animal, we might presume, has the capacity for rational thought, none has free will either. As a result, possibilities for biological evolution of free will have never previously come up. Instead the focus has always been on how free will works in humans exclusively.

Descartes and subsequently Newton, in their mathematical descriptions of the universe, precipitated a crisis for any easy acceptance of human free will. Their approach suggested that the universe has a unique diachronic pattern. Philosophers often discuss this pattern as sequences of cause and effect. The mathematics however represents each event (for instance, the rate of change in the movement or location of any object) as an analytical function of its instantaneous context. In the twentieth century, the relativistic disposition of mass and energy and the probabilities of quantal transitions were refinements of these functions. At the scale of human behavior, however, the history of the world is uniquely determined with exceedingly high probability. The past included no alternatives and the future is predictable. Everything that has happened and everything that will happen can in principle be calculated from the present. Choices among alternatives do not occur.

As an apparent confirmation of this strict determinism, neurobiologists have argued in recent years that their results also exclude free will. The pertinent results are (1) failure to locate an area in the brain (a “module”) where neural activity is associated with “choice” or “consciousness” and (2) evidence that initiating an action precedes reporting awareness of the action by a few seconds or fractions of a second. Harris, for instance, concludes that freedom from determinism is therefore an illusion. Nevertheless, recognizing this illusion can make us more sympathetic toward less fortunate people. This recognition of course would also have to be deterministic (Harris 2012).

These neurobiological results, however, do not justify any strong conclusion about whether or not brains make decisions. It is true that many activities of brains are more or less narrowly localized. Examples include primary sensory analysis, final motor control, memory, emotions, producing and understanding language (or species-specific songs in the case of some birds), and others. Yet there is no reason to think that some activities of the brain might, in contrast, be distributed widely. A capacity to make decisions might, for instance, depend on interactions of relatively few widely dispersed neurons, and, in addition, similar interactions might recur throughout large portions of the brain. Furthermore, the observation that reporting a movement follows initiating it indicates only that the act of reporting takes longer than the act of moving. When and where the pertinent decisions occur still elude us.

The invalidation of free will, which strict determinism entails, has deep ramifications for moral and legal judgments and for our sense of ourselves as individuals. Reluctant to accept these consequences of strict determinism, some philosophers such as Hume have thrown up their hands and accepted that, whatever else might be true, it is intuitively obvious that people routinely make practical choices. A similar position supposes that there are alternative worlds that a mind can choose. Others have proposed various cracks in determinism that could allow choices to seep into our behavior. In other words, someway that choice can originate actions, de novo, despite causes or analytical continuity. The usual procedure is to suppose that each person includes a supernatural component, beyond rational explanation, a soul, a “ghost in the machine,” a “categorical imperative,” a “Dasein,” a self-consciousness, and a “self-forming act,” something that makes decisions. Kane has proposed that some inherently random component of rationality underlies freedom of action. Dennett and Gazzaniga propose that the source of freedom is the human "social arena" in which interactions with other humans provide and expect reasons for actions (Kane 1996; Dennett 2002; Baer et al. 2008; Gazzaniga 2011).

These recent approaches do not differ much from earlier ones, back to Descartes and Hume. Adventitious internal randomness (quantum or otherwise) might explain erratic behavior but not rational choice. On the other hand, an intrinsic imperative for rational or moral choice simply displaces the origin of rational choice without explanation. Social arenas for human development can in many cases result in rationality and individual morality, but in other cases in rationalization and collective delusion, all strictly determined by context. To argue that humans can prejudicially accept the desirable alternatives, over the undesirable ones, assumes once again an unexplained moral imperative. Despite these shortcomings, attempts to understand freedom of choice have had the merit of emphasizing the two questions that must be addressed: (1) what is the source of unpredictability that provides an opportunity for choice and (2) what is the nature of decision.

Unpredictability And Decisions In Noise

A recent mathematical analysis of the evolution of communication by natural selection in the presence of noise reveals unexpected explanations for the unpredictability confronting an organism and for the organism's decisions. It provides an evolutionary context for investigating choice by supposing that all living animals (even plants) face these unpredictabilities and must make these decisions, each in its own way and within its own capabilities. Furthermore, it explains why the universe is deterministic but, for all organisms, the future nevertheless remains unpredictable. The crucial novelty of this approach is the inclusion of noise in an analysis of optimal performance in communication and perception. Noise here is anything that contributes to errors in reception of signals or in perceptions of sensations; errors in turn are responses disadvantageous to the receiver or perceiver in question; and a response can be overt or, in the case of a memory, covert (Wiley 2015, 2017, also see Chapter 2).

The first result of this analysis derives from signal detection theory. In the presence of noise, every receiver of any signal or perceiver of any sensation is in a double bind. Noise produces the possibility of errors, responses to signals or sensations that have net disadvantages. There are two possible kinds of errors in noise, false alarm and missed detection (errors of commission and omission). Regardless of the criteria for recognizing relevant signals or veridical sensations, it is not possible to decrease the probability of one kind of error without increasing the probability of the other.

The second result of this analysis derives from decision theory. The utility (net advantage or disadvantage) of a receiver's criterion for response depends on the intensity of the signaler's signal, and the utility of a signaler's signal depends on the stringency of the receiver's criterion for a response. Consequently, signalers and receivers evolve jointly to a mutual optimum, a Nash equilibrium at which each party does the best it can provided the other party does likewise.

The result scales to the level of noise. One prevalent determinant of the level of noise, for instance, is the distance between signalers and receivers. At close range optimal communication consists of quiet signals and lenient criteria for response. At long range the optimum consists of intense signals and stringent criteria. In all cases residual noise persists. The evolution of communication cannot escape noise. It is inevitable.

This analysis makes it clear that decisions are ubiquitous for receivers and perceivers. Every time a receiver checks its sensors, it must decide whether or not a response (or which response) is justified. Nervous systems of all types must analyze sensations and coordinate movements, but between sensation and movement, nervous systems are primarily decision-making organs.

Noise in reception or perception is not completely predictable simply because nervous systems are not complex enough to compute the dynamics of the universe. The number of neurons in a human brain is vast, but not so incomprehensibly vast as the number of interacting particles in its environment. For a brain, indeed for any practical machine, the universe is under-specified. The universe is determined but, for any brain, it remains partly unpredictable. As a result, humans remain notoriously incompetent in predicting all the consequences even of their own actions.

Unpredictability of signals and sensations thus requires decisions. These decisions, like all else in the universe, are evidently determined. Each individual's brain, like all components of every living organism, is influenced by the genes it carries and by the environment in which it lives and develops throughout its life. If a brain is complex enough to think (to use language in an internal dialogue), presumably it can weigh evidence for adopting more lenient or more stringent criteria for responses to any kind of signal or sensation. Whatever a brain thinks, as influenced by its genes and environment, is presumably determined, as are all other parts of the universe.

Nevertheless, a human brain cannot completely predict another comparable brain's activity. Explaining all interactions at any level of complexity would require a superordinate level of complexity. The collaboration among the brains of multiple people and workings of multiple other machines is making progress in understanding the general principles of how brains work. Who knows what superordinate complexity of thought might develop in the future. It remains unlikely, however, that a brain will ever completely understand and predict its own activity. Until that time, humans will continue to make decisions when faced with under-specified situations in the course of communication and perception in a noisy world.

An example of the interaction of predictability and decision in a deterministic universe is provided by chess. In this case the rules of the game are deterministic. They specify the possible moves and interactions of the pieces, in effect their "cause and effect" relationships. The number of possible sequences of moves in a game is unimaginably huge, but the number is finite (at least in versions excluding endless repetition of moves). Although the game has not been solved numerically, it is possible that there exists only one sequence of optimal moves. Humans nevertheless cannot predict this game from the outset – otherwise it would hardly be the challenge it is. Humans play the game by making decisions based on incomplete foresight. On the other hand, appropriately programmed computers can learn, indeed can teach themselves, to play with greater foresight than a human. Such a computer playing against itself or against another comparable (or more complex) computer would presumably often play the same sequence of moves.

A human cannot achieve this predictability because, with its more limited foresight, it cannot predict winning moves of a computer. A human's decisions are nevertheless determined. For a computer to predict every decision by a human, it would require even greater complexity. Such a machine would have to learn the relevant parameters of a human's brain and its context. With anything short of such complexity, a machine must make decisions based on its own formidable foresight and its human opponents' unpredictability.

The under-specified complexity of human behavior that humans and machines fail to predict is the same as noise, errors in reception and perception, as discussed above. This complexity is determined by physical laws. Nevertheless, for any particular brain or machine, if the parameters of this complexity are incompletely known, then the resulting unpredictability (noise) requires decisions.

In such a noisy world, we can legitimately judge competence at chess based on the decisions a player makes. In a similar way, we can judge moral competence based on a person's decisions in other situations. These judgments are our own decisions, our own responses to perceptions of other people in a noisy context, under-specified and thus partially unpredictable. Holding a person accountable for decisions in particular situations might require an additional decision on our part. We might require, for instance, a supplementary judgment of the person's competence for rational thought. These thoughts would also be determined, as discussed above. Nevertheless, our only evidence for another person's thoughts comes from our noisy perceptions of that person's responses. For nonhuman animals, for machines, and in some situations for humans, including ourselves, we might withhold such judgments of rationality and accountability. All of these judgments, it is important to realize, are decisions in under-specified situations, in other words, in response to noisy perceptions. They are not an indefinite regression of determinism, just responses to pervasive unpredictability in a noisy world. Noise affects everything we do.

As for the evolution of free will, this analysis of communication and perception in noise opens the possibility for comparative studies of decision-making and prediction. Indeed the fields of comparative psychology and ethology, as well as neurobiology, have made important progress. More could be done by including noise as well as signals in the comparative study of behavior or brains. The effects of noise only become apparent in the real, unpredictable world where a brain never knows as much as it would like about the source of the next sensation. It is important to investigate communication and perception in situations with multiple signals and receptors and variable sensations.

References

Baer, J., Kaufman, J. C., & Baumeister, R. F. (Eds.). (2008). *Are We Free?: Psychology And Free Will*. Oxford: Oxford University Press.

Dennett, D. C. (2002). *Freedom Evolves*. New York: Vanguard.

Gazzaniga, M. (2011). *Who's in Charge?: Free Will and the Science of the Brain*. New York: Ecco.

Harris, S. (2012). Free will. New York: Free Press.

Kane, R. (1996). *The Significance of Free Will*. Oxford: Oxford University Press

Wiley, R. H. (2015). *Noise matters: The Evolution of Communication*. Cambridge, MA: Harvard University Press.

Wiley, R. H. (2017). How noise determines the evolution of communication. Animal Behavior, 124, 307–313 [see Chapter 5].

Chapter 5

How Noise Determines the Evolution of Communication

Introduction

Following Darwin's (1972) detailed argument that animals' displays, or ‘expressions’, served for conspecific communication, almost half a century elapsed before the idea took hold among field biologists (Huxley, 1914). On first investigation, these displays seemed to be whimsical. Although Darwin had suggested his ‘Principle of Antithesis,’ according to which expressions with opposite meanings often had contrasting forms, there was scant suggestion that signals evolved to fit environmental situations. They even seemed to provide direct access to the phylogeny of species, without contamination by environmental adaptations (Heinroth, 1911, pp. 598-702; Lorenz, 1941).

This view was first shaken by Peter Marler's (1955, 1957) studies of the species distinctiveness of birds' vocalizations. He emphasized that although species specificity had advantages in some circumstances, such as territorial advertisement, it had disadvantages in other situations, such as vigilance for predators by flocks of mixed species. Furthermore, he argued that alarm calls in the latter situation had converged on sounds that were especially effective in hindering localization by predators. The time seemed right for reconsidering the importance of adaptations in animal's signals. The crucial advance came when Eugene Morton's (1975) pioneering studies revealed that birds' songs included adaptations to improve transmission through their respective habitats. Since then reports of adaptations in animals' signals have multiplied steadily. Attention has been given especially to adaptations that reduce attenuation, degradation, and effects of background environmental noise. Recently, reports have focused on human activities as widespread sources of environmental noise. Noise is now recognized to have manifold consequences for the evolution of communication. Even the residual randomness in organisms' nervous systems produces noise in communication.

Nevertheless, the crucial characteristic of noise with deep implications for the evolution of communication is still not generally appreciated. Noise, as Shannon (1948a, 1948b) originally realized, is best measured by receivers' errors. These errors are often thought just to introduce additional variance in responses to signals. As a result, adaptations to noise are assumed to consist of adjustments by signalers to minimize this extra variance. Although noise must often increase the variance of responses, it has even wider significance for the evolution of communication, because noise produces unavoidable trade-offs for any receiver. A receiver cannot maximize its performance in the presence of noise; it can only optimize these trade-offs. Furthermore, not only does the optimal behavior of receivers depend on the behavior of signalers, but the optimal behavior of signalers also depends on the behavior of receivers. Neither the evolution of signalers nor the evolution of receivers can be convincingly explained without taking into account the full consequences of noise.

Previous efforts to explain the evolution of signaling include those that emphasize the evolution of honesty (Enquist, Plane, & Roäed, 1985; Getty, 1998; Grafen, 1990; Hurd, 1995; Johnstone, 1995; Maynard Smith, 1991; Maynard Smith & Harper, 2003; Sza mado & Penn, 2015; Zahavi, 1977; Zahavi & Zahavi, 1997), those that focus on the dynamics of mate choice (for instance, Kirkpatrick, 1982; Lande, 1981; Servedio, 2011) and those that focus of the evolution of stable cooperative interactions (for instance, Scott-Phillips, Blythe, Gardner, & West, 2012; Scott-Phillips & Kirby, 2013). Some previous analyses include the effects of noise as additional variance in responses (Johnstone, 1994) and even emphasize the consequences of the receiver's trade-offs in noise (Johnstone, 1998; Wiley, 1994), but none includes these trade-offs in combination with full interdependence of the receiver's and signaler's performances.

A recent effort to understand the interaction of receiver and signaler in noise has produced some unexpected results (Wiley, 2013a, 2013b, 2015). Some long-standing problems, such as conditions for the evolution of honesty and for evolutionarily stable signaling, appear in an entirely new light. The evolution of mate choice takes on a new dimension. Furthermore, it also becomes apparent that some critical features of communication have so far not received much, or any, investigation. The mathematical analysis of the optimal behavior for receiver and signaler in noise has been described elsewhere (Wiley, 2013a, 2015). This essay instead isolates a dozen principles, or distinctive predictions, of the evolution of communication in noise. They reveal that noise is an essential factor in the evolution of all communication.

Twelve Predictions

This essay identifies a dozen predictions from this recent analysis of the evolution of communication in the presence of noise (Wiley 2013a, 2015). Perhaps it is best to list these predictions at the outset and then to focus on each in turn.

First of all, (1) noise creates an unavoidable trade-off between two kinds of error by receivers. Furthermore, (2) a receiver's optimal criterion for response depends on the level of signals and (3) a signaler's optimal level of signaling depends on the receiver's criterion.

As a result, (4) communication in noise can evolve to a joint optimum. (5) Communication at a joint optimum is honest on average. (6) Joint optima for communication in noise do not eliminate noise. (7) Many parameters of communication in noise remain poorly studied. (8) Noise leads to strong predictions for the evolution of exaggeration and thresholds. (9) Signals for advertising and for warning are contrasts in probable costs of errors. (10) The evolution of new signals and responses encounters a hurdle. (11) New signals and responses can evolve by exploitation. (12) Joint evolution of signalers and receivers has a predictable direction.

The first of these twelve predictions is the most fundamental.

(1) Noise Creates an Unavoidable Trade-Off between Two Kinds of Error by Receivers

In the presence of noise, there are exactly four possible outcomes each time a receiver makes a decision to respond or not: correct detection, correct rejection, false alarm and missed detection. These four possibilities are the logical combinations of two possible external conditions (noise only or noise plus signal) and two possible decisions by a receiver (respond or not). Two of the four are errors: false alarm and missed detection. In an analysis of the evolution of communication, these two would result in lower survival or reproduction. These two kinds of error are conceptually the same as type I and type II errors in analyses of statistical significance, or errors of commission and errors of omission. The probabilities of the four possible outcomes define a receiver's performance in any particular situation, a situation thoroughly analyzed by signal detection theory (Green & Swets, 1966; Macmillan, 2002; Macmillan & Creelman, 2005).

These four outcomes are also a direct consequence of the defining feature of communication -- responses (changes in behavior) by one party (a receiver) to signals by another party (a signaler). A signal in this context is any pattern of energy and matter that can evoke a response without providing all of the power for the response (Wiley, 1994, 2006, 2013c). As a consequence, a receiver must make the decision to respond. To do so, it must include sensors (to detect impinging energy and matter), gates (switches to determine which inputs elicit a response), and amplifiers (to provide the additional power for the response). A receiver's gate for a particular response might take the form of a threshold (a minimal level of activation of the sensor) or a filter (an optimal level of activation) e or complex combinations of these two to produce a cognitive criterion for response.

The four possible outcomes each time a receiver checks its sensor are an exhaustive and mutually exclusive categorization of possibilities. Whenever a receiver's sensor cannot absolutely eliminate noise, these four possibilities recur. Furthermore, the two kinds of error cannot be simultaneously minimized. Adjusting a threshold or filter to reduce one inevitably augments the other (Wiley, 1994, 2006). False alarms and missed detections are therefore an inevitable trade-off for any receiver in noise. Noise does not just create extra variance in responses; it puts every receiver in a double bind.

(2) A Receiver's Optimal Criterion for Response Depends on the Level of Signals

Because of the inevitable trade-off between two kinds of errors, a receiver cannot minimize its errors overall; the best it can do is to choose a criterion for response that optimizes the trade-off. The criterion for an evolutionary optimum depends on (1) the probabilities of the four possible outcomes and (2) the consequences of each outcome for the receiver's survival and reproduction (the evolutionary payoff for each outcome). The probability of a correct detection, for instance, is a product of the probability that a signal actually occurs at the moment a receiver checks its sensor and the probability that the receiver responds in this situation. The probability that the receiver responds when a signal occurs depends in turn on its criterion for responses (the location of its threshold, for instance) and on the level of the signal in relation to any noise (the signal/noise ratio). In general, the probability of each of the four possible outcomes depends on (1) the probability that a signal occurs, (2) the receiver's criterion for response and (3) the level of the signal in relation to noise. A linear combination of these probabilities and payoffs for the four possible outcomes specifies the utility of a receiver's criterion for response (Wiley, 1994, 2013a, 2015). This approach is the basis of decision theory (van Neumann & Morgenstern, 1953).

Maximizing this utility depends on the trade-offs between the two possible errors and between the two possible correct responses. It also depends on the level of the signal in relation to the noise (the signal/noise ratio). Consequently, the receiver's optimal criterion for response depends in part on the level of signal produced by the signaler.

(3) A Signaler's Optimal Level of Signaling Depends on the Receiver's Criterion

Often, perhaps always, a higher level of signaling (greater intensity, size or saturation, or in general greater 'exaggeration') comes with costs, as a result of greater expenditure of energy, commitment of time, opportunities lost, or exposure to inappropriate receivers (such as predators, parasites or competitors). There have previously been two lessons drawn from these costs of signaling: (1) costs are necessary for the evolution of honest signaling (sometimes with a provision that the costs must be 'wanton' or 'excessive') (Maynard Smith & Harper, 2003; Zahavi & Zahavi, 1997); and (2) increasing costs multiplied by increasing benefits can produce evolutionarily stable signals, which in turn are honest (Getty, 1998; Nur & Hasson, 1984; Wiley, 2000, 2015).

It is easy to show that combinations of benefits and costs can produce equilibrial levels of signaling (including signals for advertisement and for solicitation; see Appendix and Wiley, 2000, 2015). These treatments however ignore the interdependent evolution of the signaler and receiver. The benefit for the signaler comes from responses (correct detections) by appropriate receivers, and the probability of these responses depends on the optimal criterion for response by these receivers.

Thus the optimal level of signals by a signaler cannot be determined without reference to the performance of the appropriate receivers. At the same time, as introduced in the preceding section, a receiver's optimal criterion cannot be determined without reference to the signaler's level of signaling. It is not possible for either party to optimize its behavior on the basis of fixed costs and benefits. So no argument that honesty (or any other feature of communication) depends only on a signaler's costs can be complete. Instead, the only way to understand the evolution of communication in the presence of noise is to consider the possibility of a joint optimum, one at which the receiver's criterion is optimal provided the signaler's exaggeration is optimal and vice versa.

Notice that the preceding definition of a signal differs from previous ones especially in lacking any qualification that signals (as opposed to cues) must have evolved for the purpose of communication. This qualification, which has the unfortunate consequence of making the definition of signals and communication circular, is unnecessary (Wiley, 2013c, 2015). On the other hand, it is apparent that when signalers, as well as receivers, are living organisms, then each can evolve in relation to the other.

(4) Communication in Noise Can Evolve to a Joint Optimum

By proposing utility functions for both the receiver's threshold for response and for the signaler's level of exaggeration, it is possible to derive optimal thresholds for every level of exaggeration and, conversely, optimal exaggerations for every level of threshold. It then becomes possible to search for points of coincidence between these optima for threshold and for exaggeration (Wiley, 2013a, 2015).

The result depends on the payoffs for the possible outcomes of a receiver's decisions and on the cost of exaggeration and the benefit to a signaler from a receiver's correct detection of a signal. The result also depends on the frequency of signals, both overall (which affects the signaler's overall cost) and at times when receivers are monitoring their sensors (which affects the probabilities of a receiver's four possible outcomes). It turns out that communication in noise, with reasonable conjectures for these parameters, often leads to evolutionarily stable levels of a receiver's threshold and a signaler's exaggeration (Wiley, 2013a, 2015).

Calculation of the adaptive landscapes around these joint optima show that they are Nash equilibria: joint optima at which each party does the best it can provided the other does also. On the other hand, there are often loci in the adaptive landscapes where evolution can diverge, either towards a joint optimum or towards a collapse of communication. Such a collapse occurs when the optimal exaggeration = 0, indicating no signal, or the optimal threshold = 0, indicating no discrimination and thus no association of responses with signals.

Evolution through the adaptive landscape defined by levels of receivers' thresholds and levels of signalers' exaggeration is a process that involves continual adjustments of both thresholds for response and exaggeration of signals. The exact course of evolution towards a joint optimum depends on the starting conditions and on the payoffs, costs, benefits and probabilities already mentioned, but in no case does either receiver or signaler evolve in relation to a fixed level of performance by the other party.

(5) Communication at a Joint Optimum Is Honest on Average

Communication is honest because at these joint optima receivers benefit (utility > 0). It has been recognized previously that receivers must benefit ‘on average’ or overall for communication to evolve, otherwise it would not pay for receivers to attend to signals (Grafen, 1990; Guilford & Dawkins, 1991). Any costs of deception or manipulation of receivers (or exploitation of signalers) must be more than balanced by benefits. Nevertheless, the emphasis has often been placed on the costs for the signaler (Lachmann, Számadó, & Bergstrom, 2001; Számadó, 2011). In contrast, the evolution of communication in noise shows that receivers and signalers evolve to a joint optimum at which each party benefits on average and each optimizes its behavior provided the other does also. Despite adaptations to reduce the effects of noise, possibilities for deception, manipulation and exploitation persist. So do benefits on average for both parties. Communication in noise predicts the evolution of both honesty on average and residual manipulation.

(6) Joint Optima for Communication in Noise Does Not Eliminate Noise

The joint optima for receiver and signaler in noise never reach a level at which noise is eliminated. Both parties face diminishing benefits and augmenting costs as thresholds and exaggeration rise. As a result communication evolves to optimize performance in noise by reducing the consequences of noise but not by eliminating noise. In high levels of noise, communication can evolve high thresholds and high exaggeration, yet always retain the possibility of errors by receivers and signals without responses. Likewise, in low levels of noise, communication can evolve low thresholds and low exaggeration, yet still retain possibilities for errors and frustration. In both situations, the optimal performances of receiver and signaler scale to the level of noise.

(7) Many Parameters of Communication in Noise Remain Poorly Studied

This new analysis indicates that most of the parameters that influence the evolution of communication in noise are not well known. For instance, the probabilities of the four possible outcomes and the relative frequencies of signals have received little or no attention. Some of the costs, benefits and payoffs, in particular the payoffs for false alarms or missed detections, are also often neglected. On the other hand, the costs of signals and the benefits of correct detections are better known.

There are many reports of the costs of signals (displays), well summarized by Bill Searcy and Steve Nowicki (2005). Møller's (1994) pioneering studies of the costs of tail streamers in barn swallows *Hirundo rustica* provided a model for demonstrating that individuals of higher quality could produce larger displays with less overall cost than could individuals of lower quality. Yet this study did not estimate the marginal cost of exaggeration. The available data suggest that the costs are not linear with exaggeration in this case (see Wiley, 2015). There seems to be no reason why future studies of this or other displays could not estimate marginal costs of exaggeration as well as intrinsic quality of signalers (for instance, by extrapolation to survival in the absence of any signaling). Overall costs of a signal do not alone provide a way to deduce either of these other parameters.

Benefits of responding to optimal signals have also received attention, especially in the case of choosing an optimal as opposed to a suboptimal mate (again Searcy & Nowicki, 2005). These comparisons, however, have never considered the payoffs for all four possible outcomes for a receiver (or at least the relative payoffs for three of them in comparison to the fourth). There have also been estimates of the risks of signals for predation and parasitism. Studies of mimicry have estimated nearly all the costs and benefits of signals, and even their relative frequencies (Kikuchi & Pfennig, 2013), but we know less about the payoffs for the four possible outcomes for receivers.

These examples are enough to suggest that all of the parameters relevant to understanding the evolution of communication in noise can be estimated in natural conditions e like those in which the signals and responses evolved. Because, as emphasized above, escape from noise is not expected, a full understanding of the evolution of communication must include more attention to these neglected or ignored parameters.

(8) Noise Leads to Strong Predictions for the Evolution of Exaggeration and Thresholds

This new analysis permits an analysis of the sensitivity of the evolution of communication in noise to variation in each of the parameters. An important caveat is that a predicted effect of changing any one parameter requires that all other parameters remain constant. Foremost among these analyses is the prediction that (1) high marginal costs of exaggeration (the cost of each unit of exaggeration) result in lower levels of exaggeration and lower thresholds for response. Perhaps contrary to current expectations, exaggerated signals are predicted to have low marginal costs (all else equal). Furthermore, the payoffs for each of the four possible outcomes of a receiver's decision to respond or not affect both its optimal threshold and the signaler's optimal level of exaggeration. As a result, there are two more predictions to make. (2) Higher costs (lower payoffs) of false alarms for receivers lead to higher levels of exaggeration by signalers. (3) Higher benefits of correct detections (higher payoffs) for receivers lead to lower levels of exaggeration by signalers. Lower exaggeration also results from higher costs (lower payoffs) of missed detections.

Furthermore, higher relative frequencies of signaling result in both higher overall costs for signalers and lower thresholds for receivers (because correct detections become inherently more likely than false alarms). For both reasons, (4) higher relative frequencies of signaling lead to lower levels of exaggeration. In the limit, when signals always occur whenever a receiver checks its threshold, it no longer pays for the receiver to bother; instead it pays to respond at any time, and the evolution of communication collapses.

The parameters that are the basis for these predictions e the marginal cost of signals, the payoff for a false alarm in comparison to that for a correct detections and the relative frequency of a signal e are all poorly known. Yet they have strong influences on the predicted evolution of communication in noise.

(9) Signals for Advertising and for Warning Are Contrasts in Probable Costs of Errors

Advertising for mates is the classical case for exaggerated signals. In this case the receiver (an individual of the choosy sex) encounters signals (displays by high-quality potential mates) as well as noise (displays by low-quality potential mates). The choosy sex is often supposed to have coy behavior (frequent failures to respond to high-quality prospects). In other words, receivers (choosers) accept many missed detections (passing optimal mates). They would thereby minimize false alarms (accepting suboptimal mates). Coy behavior thus corresponds to a high threshold for response e adaptive fastidiousness (Wiley, 1994).

In contrast, signals for warning presumably have the converse relationship between the costs of false alarms and missed detections. Receivers that miss a warning risk exposure to a dangerous predator. A false alarm, by responding for instance to a deceptive warning signal, might often entail only a brief interruption of feeding or courtship. A high cost for missed detections, in comparison to false alarms, would result in a low threshold for response. It would be manifest as jumpy receivers that often responded to deceptive signals e adaptive gullability (Wiley, 1994).

High thresholds for coy receivers choosing mates and low thresholds for jumpy receivers attending to warnings suggest that signals for advertisement should have high exaggeration and those for warning should have low exaggeration (Wiley, 1994). Yet analyses of the evolution of noisy communication with some hypothetical payoffs for the four possible outcomes of a receiver's decisions to respond have not confirmed that the contrast in payoffs for false alarms and missed detections produce the expected contrast in exaggeration of signals (Wiley, 2015, contra Wiley, 2013a). Instead a contrast in frequency of signals counteracts the contrast in costs of errors so that both warning and advertising signals are expected to evolve exaggeration. Warning signals in this analysis evolve high exaggeration, despite high costs of missed detection, because they are relatively infrequent; advertising signals evolve high exaggeration, despite high frequency, because false alarms are relatively costly.

(10) Evolution of New Signals and Responses Encounters a Hurdle

The adaptive landscapes for the evolution of signaler and receiver in noise illustrate an intuitive conclusion about the origin of new signals. New signals cannot evolve in the absence of appropriate responses; and responses cannot evolve in the absence of suitable signals. The evolution of new signals and responses corresponds to initial conditions with high thresholds and low exaggeration. No response corresponds to an infinite threshold, so an incipient response would correspond to a high threshold. No signal is zero exaggeration, so an incipient one would have low exaggeration. The corresponding quadrant of the adaptive landscapes in noise epitomizes this problematic condition for the evolution of new signals or new responses ab initio. The recent analysis shows that selection gradients in these conditions move signalers and receivers towards a collapse of communication, towards no exaggeration of signals and no lower thresholds (Wiley, 2013a, 2015). This analysis is more precise than the intuitive adage, because it shows that the collapse of incipient communication is a result of signalers and receivers jointly optimizing their behavior.

A similar situation has long been recognized for the evolution of signals for mate attraction in quantitative genetic models of sexual selection. The strength or prevalence of females' preferences (in the prevalent situation with female choice) must exceed a threshold before the evolution of males' traits begins to accelerate. The new analysis of communication in noise shows, in a quantitative phenotypic analysis, that a similar condition applies to the evolution of all communication. The evolution of communication *ab initio* must cross a hurdle.

(11) New Signals And Responses Can Evolve By Exploitation

There are two ways that might lower or eliminate this hurdle. Appropriate terms for these two options are sensory exploitation by receivers and incipient signalers and motor exploitation by signalers and incipient receivers. ‘Exploitation’ here is meant to suggest that both signalers and receivers can jointly take advantage of their particular features, not that one party takes advantage of the other (Ryan, 1990, suggests instead that signalers exploit receivers). Sensory exploitation would occur when some individuals, as a result of adaptations having nothing to do with communication or at least with the newly evolving form of communication, already have responses to particular sensory input. As a result, an incipient (mutant) signal that also evoked that response might encounter initial conditions for evolution outside the problematic quadrant of adaptive fields for joint evolution of signaler and receiver in noise.

The exact initial conditions would depend on the payoffs for the receivers and the costs and benefits of exaggeration for the signalers and the relative frequency of the new signal. If the initial conditions avoided the collapse of communication in the problematic quadrant, then joint evolution of signals and responses would proceed towards the appropriate joint optimum. Both parties would evolve in this process, so there would be no implication that one party was taking advantage of the other. Both might however benefit from the circumstances that allowed a new system of signaling and responding to jump the hurdle for their evolution ab initio. Sensory exploitation in this sense is related to previous proposals, but without the implication that one party takes advantage of the other.

The alternative way to jump the hurdle *ab initio* is motor exploitation. In this case an action (a movement or a synthesis of a structure or molecule) might already exist as an adaptation unrelated to communication (or the novel form of communication). For instance, consider the suggestion by early ethologists that comfort movements or 'displacement activities' often provide the initial condition for the evolution of new displays (Tinbergen, 1940, 1959). Such actions might also indicate something about the performer that could make it advantageous for another individual to respond. Just as with sensory exploitation, this situation could provide initial conditions for the evolution of a new system of signal and response that lay outside the problematic corner of the adaptive landscape for evolution in noise. As before, the initial conditions would then result in joint evolution of signaling and responding towards a joint optimum for signalers and receivers. It would not be a case of one party exploiting the other. Instead both parties would exploit their complementary features that permit the evolution of a mutually beneficial signal and response. Sensory and motor exploitation are examples of cooptation in evolution.

(12) Joint Evolution of Signalers and Receivers Has a Predictable Direction

The evolution of communication in noise evolves towards a joint optimum of thresholds for response and exaggeration of signals. Exaggeration in this context consists of adaptations that increase the detectability of the signals. Detectability (or discriminability) of signals depends partly on the properties of the receiver's sensors including their thresholds or filtering of input and their levels of intrinsic noise. Detectability is also influenced by the properties of signals. The relevant properties are often summarized as the signal/ noise ratio.

Contrast between signal and noise is directly related to this ratio. Contrast results from the intensity of a signal, especially in those features with low intensity in external noise. Saturation (concentration of energy or matter in particular features of the signal) also contributes to contrast provided the appropriate receivers' sensors can differentiate these features. Examples include a concentration of acoustic energy in a particular frequency of sound at any instant, as do many birds' songs, or concentration of optic energy in a particular wavelength at any point (~1/frequency), as do iridescent colors. The evolution of communication in noise predicts that signals evolve optimal levels of exaggeration in the specific sense of contrast with environmental noise (whether from nonbiological, heterospecific or conspecific sources).

Predictability also contributes to the detectability of signals. Almost any prior knowledge about (or prior experience with) parameters of a signal (including its timing and location) makes it more detectable. For instance, an alerting signal, easily detectable but information-sparse, is one way to increase the predictability of a contiguous signal that is information-dense (Wiley & Richards, 1982). Redundancy, as well known, also enhances the detectability of signals. Contrast, predictability and redundancy are the features of ‘ritualized’ signals, which early ethologists proposed had evolved to facilitate communication. They had, however, not emphasized the particular advantage of these features of signals for communication in noise.

Notice that sexual selection also predicts progressive evolution of signals in accordance with one sex's preferences. It provides quantitative predictions for the dynamics of this joint evolution. On the other hand, it does not provide predictions about the direction of evolution. The theory of communication in noise, summarized here, has complementary advantages and disadvantages. It predicts the direction (and ultimate equilibrial optimum) but not the dynamics of the joint evolution of signals and responses.

Conclusion

These dozen points about the evolution of communication in noise suggest that the evolutionarily optimal properties of communication should be understandable in detail in any particular situation. Noise is critical for this understanding, because the most basic prediction from analyzing the evolution of communication in noise is that communication does not evolve to eliminate noise. Noise is thus expected to be a persistent feature of all communication. These dozen points are conclusions and extrapolations from a full mathematical analysis presented elsewhere (Wiley, 2013a, 2015). Perhaps the biggest lesson from this analysis is the number of parameters that must be understood to explain the evolution of communication in noise. There are 10 of these parameters (Wiley, 2015, p. 185). Most have received little attention by students of animal (or any other form of) communication. Those most neglected do not seem inherently more difficult to measure than those that have already received some study. Until all of these parameters get some attention, the predictions from the mathematical analysis will remain untested. Only by accounting for noise will it be possible to understand the evolution of communication.

Many students and colleagues contributed to developing and critiquing this analysis of communication in noise (see Wiley, 2015). Nevertheless, I would like to acknowledge with special emphasis the impetus to this study that came from Peter Marler. He was unrivaled in pioneering diverse approaches to the study of communication, in promoting the development of these approaches in his students and colleagues, and in explicating the issues with exceptional clarity. Without the scientific family he nurtured around the study of communication, the belated efforts presented here would never have been transmitted.

Appendices

(1) Optimal Levels of Exaggeration for Advertisement

In this case the costs of signaling are lower survival and the benefits are higher reproduction. Such a situation might apply to males advertising for mates. Suppose that (1) signals have costs and benefits that increase with exaggeration, (2) costs can be expressed as decreased survival, (3) benefits can be expressed as increased reproduction, and (4) the influence of natural selection on the spread of genes can be approximated by survival reproduction of phenotypes associated with those genes. If the cost increases (survival decreases) linearly with exaggeration and the benefit (reproduction) increases linearly with exaggeration, then taking the derivative of the product of these two functions with respect to exaggeration and setting the derivative to 0 shows that maximal survival benefit occurs when

$$e^* = -i/2m - o/2g$$

In this expression, e^* is the optimal level of exaggeration, i is intrinsic survival (in the absence of any signaling), m is the marginal cost of exaggeration (the negative slope of survival as a function of exaggeration), o is the offset for reproduction ($o > 0$ indicates a residual level of reproduction without signaling and $o < 0$ indicates that exaggeration must reach some level before any reproduction occurs), and g is the marginal gain in reproduction (the positive slope of reproduction as a function of exaggeration). The second derivative of survival reproduction with respect to exaggeration confirms that this optimal level of exaggeration is indeed a maximum, provided $m < 0$ and $g > 0$. The above expression shows that signalers with higher intrinsic survival or lower marginal cost of exaggeration have higher optimal levels of exaggeration. Signalers with higher quality by either of these two measures should thus have more exaggerated signals, provided all signalers produce optimal levels of exaggeration. It would not pay for low-quality signalers to produce as much signal as high-quality signalers, even though their overall survival reproduction is less.

(2) Optimal Levels of Exaggeration for Solicitation

In this case signalers balance one source of mortality against another. Young begging for food from parents might fit this situation, when begging decreases the risk of starvation but increases the risk of predation (for instance, if begging attracts predators or parasites). If the first risk exceeds the second, then begging pays and natural selection would favor signals that minimize the overall risk of death. A complexity arises because the probability of starvation $P(\mathrm{s})$ is not independent of the probability of predation $P(\mathrm{p})$, so the overall mortality equals $P(\mathrm{s}) + P(\mathrm{p}) - P(\mathrm{s})P(\mathrm{p})$.

Assume that the probability of starvation $P(\mathrm{s}) = S + e\ s'$, where e is the exaggeration of signals, S is the intrinsic risk of starvation without any signaling, and s' is the negative marginal risk of starvation with increasing exaggeration (negative to indicate that the chance of starvation decreases with increased begging). Analogously, assume that the probability of predation $P(\mathrm{p}) = P + e\ p'$, where P is the intrinsic risk of predation without any signaling and p' is the positive marginal risk of predation with increasing exaggeration of signal (positive to indicate that the chance of predation increases with increased exaggeration).

Expanding the equation above for overall mortality, then taking the derivative of the result with respect to e and zeroing the derivative, reveals a unique level of exaggeration that minimizes overall mortality:

$$e^{*} = \frac{1}{2}\left(\frac{1-P}{p'} + \frac{1-S}{s'}\right)$$

provided that P and $S > 0$, $p' > 0$ and $s' < 0$. The optimal exaggeration of soliciting signals thus increases as each of the four parameters, P, S, p' or s', increases. Because s' is negative by definition, an increase in s' is equivalent to a decrease in the absolute value of s'. In the case of young sharing a nest, if one is in better condition (better fed, for instance) than another, it might thereby have lower S and lower s'. In other words, it might be intrinsically less likely to starve and also its chance of starvation might decrease proportionately less for each unit of exaggeration. On the other hand, P and p' might not differ among nestlings regardless of their condition, if predators are likely to take all young once a nest is discovered. If so, the preceding equation predicts that well-fed nestlings maximize survival by begging less than other nestlings.

These calculations have not included any indirect effects of an individual's behavior on survival of relatives, such as siblings and parents. Incorporating these effects would require rephrasing the argument in terms of a change in the expected number of copies of a gene in the next generation, by adding any effects of an individual's signals on the expected survival of relatives, depreciated by their genealogical relatedness, to the individual's own expected survival. Because an individual's signals would usually decrease the expected number of relatives in the next generation (particularly when begging decreases a parent's condition or increases predation on nests), these effects would tend to reduce the optimal exaggeration of individuals' signals for begging (see Godfray, 1995).

These calculations also do not take into account the coevolution of joint optima by signaler and receiver, as summarized above.

References

Darwin, C. (1972). *The Expression of the Emotions in Man and Animals*. London, U.K.: J. Murray.

Enquist, M., Plane, E., & Roäed, J. (1985). Aggressive communication in fulmars (*Fulmarus glacialis*) competing for food. *Animal Behaviour*, 33, 1007-1020.

Getty, T. (1998). Handicap signaling: When fecundity and viability do not add up. *Animal Behaviour*, 56, 127-130.

Godfray, H. C. T. (1995). Signaling of need between parents and young: Parent-offspring conflict and sibling rivalry. *American Naturalist*, 146, 1-24.

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

Green, D. M., & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: John Wiley.

Guilford, T., & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, 42, 1-14.

Heinroth, O. (1911). Beiträge zur Biologie, namentlich Ethologie und Psychologie der Anatiden. Verhandlungen des V. 1910. Berlin: *Internationalen Ornithologenkongressen*.

Hurd, P. L. (1995). Communication in discrete actioneresponse games. *Journal of Theoretical Biology*, 174, 217-222.

Huxley, J. S. (1914). The courtship-habits of the great crested grebe (Podiceps cristatus); with an addition to the theory of sexual selection. *Proceedings of the Zoological Society of London*, 35, 491-562.

Johnstone, R. A. (1994). Honest signaling, perceptual error and the evolution of 'all-or-nothing' displays. *Proceedings of the Royal Society B: Biological Sciences*, 256, 169-175.

Johnstone, R. A. (1995). Sexual selection, honest advertisement and the handicap principle: Reviewing the evidence. *Biological Reviews*, 70, 1-65.

Johnstone, R. A. (1998). Conspiratorial whispers and conspicuous displays: Games of signal detection. *Evolution*, 52, 1554-1563.

Kikuchi, D. W., & Pfennig, D. W. (2013). Imperfect mimicry and the limits of natural selection. *Quarterly Review of Biology*, 88, 297-315.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, 36, 1-12.

Lachmann, M., Sza mado, S., & Bergstrom, C. T. (2001). Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 13189-13194.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences of the United States of America*, 78, 3721-3725.

Lorenz, K. (1941). Vergleichende Bewegungsstudien bei Anatiden. *Journal für Ornithologie*, 89, 194-294.

Macmillan, N. A. (2002). Signal detection theory. In H. E. Pashler (Ed.), *Stevens' Handbook of Experimental Psychology* (3rd ed., Vol. 4, pp. 43-90). New York, NY: Wiley.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide*. Mahwah, NJ: L. Erlbaum.

Marler, P. (1955). The characteristics of certain animal calls. *Nature*, 176, 6-8.

Marler, P. (1957). Specific distinctiveness in the communication signals of birds. *Behaviour*, 11, 13-38.

Maynard Smith, J. (1991). Honest signaling: The Philip Sidney game. *Animal Behaviour*, 42, 1034-1035.

Maynard Smith, J., & Harper, D. (2003). *Animal Signals*. Oxford, U.K.: Oxford University Press.

Møller, A. P. (1994). *Sexual Selection And The Barn Swallow*. Oxford, U.K.: Oxford University Press.

Morton, E. S. (1975). Ecological sources of selection on avian sounds. *American Naturalist*, 109, 17-34.

Neumann, J. V., & Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Nur, N., & Hasson, O. (1984). Phenotypic plasticity and the handicap principle. *Journal of Theoretical Biology*, 110, 275-297.

Ryan, M. J. (1990). Sexual selection, sensory systems and sensory exploitation. *Oxford Surveys in Evolutionary Biology*, 7, 157-195.

Scott-Phillips, T. C., Blythe, R. A., Gardner, A., & West, S. A. (2012). How do communication systems emerge? *Proceedings of the Royal Society B: Biological Sciences*, 279, 1943-1949.

Scott-Phillips, T. C., & Kirby, S. (2013). Information, influence and inference in language evolution. In U. E. Stegmann (Ed.), *Animal Communication Theory: Information and Influence* (pp. 421-438). Cambridge, U.K.: Cambridge University Press.

Searcy, W. A., & Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems*. Princeton, NJ: Princeton University Press.

Servedio, M. R. (2011). Limits to the evolution of assortative mating by female choice under restricted gene flow. *Proceedings of the Royal Society B: Biological Sciences*, 278, 179-187.

Shannon, C. E. (1948a). The mathematical theory of communication, I. *Bell System Technical Journal*, 27, 379-423.

Shannon, C. E. (1948b). The mathematical theory of communication, II. *Bell System Technical Journal*, 27, 623-656.

Számadó, S. (2011). The cost of honesty and the fallacy of the handicap principle. *Animal Behaviour*, 81, 3-10.

Számadó, S., & Penn, D. J. (2015). Why does costly signaling evolve? Challenges with testing the handicap hypothesis. *Animal Behaviour*, 110, e9-e12.

Tinbergen, N. (1940). Die Übersprungbewegung. *Zeitschrift für Tierpsychologie*, 4, 1-40.

Tinbergen, N. (1959). Comparative studies of the behavior of gulls (Laridae): A progress report. *Behaviour*, 15, 1-70.

Wiley, R. H. (1994). Errors, exaggeration, and deception in animal communication. In L. Real (Ed.), *Behavioral Mechanisms in Ecology* (pp. 157-189). Chicago, IL: University of Chicago Press.

Wiley, R. H. (2000). Sexual selection and mate choice: Trade-offs for males and females. In M. Apollonio, M. Festa-Bianchet, & D. Mainardi (Eds.), *Vertebrate Mating Systems* (pp. 8-46). Singapore: World Scientific.

Wiley, R. H. (2006). Signal detection and animal communication. *Advances in the Study of Behavior*, 36, 217-247.

Wiley, R. H. (2013a). A receiver-signaler equilibrium in the evolution of communication in noise. *Behaviour*, 150, 957-993.

Wiley, R. H. (2013b). Signal detection, noise, and the evolution of communication. In H. Brumm (Ed.), *Animal Communication and Noise* (pp. 7-30). Heidelberg, DE: Springer.

Wiley, R. H. (2013c). Communication as a transfer of information: Measurement, mechanism and meaning. In U. E. Stegmann (Ed.), *Animal Communication Theory: Information And Influence* (pp. 421-438). Cambridge, U.K.: Cambridge University Press. [see Chapter 9]

Wiley, R. H. (2015). *Noise Matters: The Evolution of Communication*. Cambridge, MA: Harvard University Press.

Wiley, R. H., & Richards, D. G. (1982). Adaptations for acoustic communication in birds: Sound transmission and signal detection. In D. E. Kroodsma, & E. H. Miller (Eds.), *Acoustic Communication in Birds* (Vol. 1, pp. 132-181). New York: Academic Press.

Zahavi, A. (1977). The cost of honesty: Further remarks on the handicap principle. *Journal of Theoretical Biology*, 67, 603-605.

Zahavi, A., & Zahavi, A. (1997). *The Handicap Principle: A Missing Piece of Darwin's Puzzle*. Oxford, U.K.: Oxford University Press.

Chapter 6

Natural Selection

Introduction

Charles Darwin described natural selection in private essays in 1842 and especially 1844 (Darwin 1909; Glik and Kohn 1996, pp. 90–96). He then drafted a large manuscript on natural selection, which he left unfinished. It was subsequently overlooked until recently (Darwin 1975). Eventually in 1858, the Linnean Society published a version of his 1844 essay in conjunction with a communication from Alfred Russell Wallace. Wallace presented some related ideas, but not natural selection as we now understand it (Bulmer 2005). Soon afterward there appeared On the *Origin of Species / By Means of Natural Selection, / or the / Preservation of Favoured Races in the Struggle for Life* (Darwin 1859), which developed the concept in detail.

Introducing natural selection on the first few pages, Darwin emphasized the importance of variation among individuals, in particular hereditary variation, and a "struggle for existence," in other words, competition, because "many more individuals of each species are born than can possibly survive". He thus reasoned, "It follows that any being, if it vary however slightly in any manner profitable to itself ... will have a better chance of surviving and thus be naturally selected. From the strong principle of inheritance, any selected variety will tend to propagate its new and modified form." He also recognized that natural selection is a means of the "coadaptation of organic beings to each other and to the physical conditions of life" (Darwin 1859, pp. 4–5).

Inherited differences in reproduction, as well as survival, can also lead to natural selection. Darwin emphasized this possibility when he proposed that differences in attracting mates or competing for them could lead to the special case of sexual selection. Attraction of mates proved to be especially controversial for Darwin's successors. It implied the evolution of behavior, not just morphology. The idea that natural selection might produce preferences for mates, particularly by females, was inconceivable to most scientists in the late nineteenth century.

Darwin's original theory thus included all of the essential elements of our current understanding of natural selection (see, for instance, Maynard Smith 1998). Natural selection is a mechanism of evolutionary adaptation that results from a combination of heritable variation among individuals and differences in their survival or reproduction correlated with this variation. Natural selection does not require individuals to change, but it does require new variants to arise occasionally through reproduction. Natural selection requires more than individual differences in survival and reproduction. It also requires heritable variation in these differences. In short, natural selection is differences in the propagation of genes in a population as a result of survival and reproduction of organisms carrying those genes. Darwin's presentation of natural selection included these elements but left many uncertainties, largely because the sciences of genetics and ecology had yet to come.

Basic Issues

When Darwin proposed natural selection, no one understood the mechanisms of heredity. Darwin himself conducted extensive experimental investigations of heredity and selection in domestic pigeons, but he was disturbed by his results. Offspring of differing parents often either combined the parental features or had intermediate features. Inheritance in this way, by blending of parental traits, could not produce adaptation by natural selection. Populations would instead converge on an overall average and then no longer change. Nevertheless it was apparent that pigeons did sometimes inherit parental features without complete blending. Within 50 years, examples of particulate inheritance and its infrequent mutations had been thoroughly verified. The inchoate field of genetics had provided Darwin's theory with the requisite mechanisms for heredity and variation, in the form of genes, each with variants (alleles).

Furthermore, it became clear that all steps in the process of natural selection – differences in survival and reproduction, heritabilities, and rates of mutation – were measurable and thus open to mathematical analysis. Within the first decades of the twentieth century, the mathematical theory of evolution by natural selection had established its basic principles. Evolution, that is, a change in the frequencies of alleles in a population of organisms, depends quantitatively on a balance between selection, mutation, and migration between populations, as well as the inherent randomness in each of these three processes. Selection is included in equations for changes in allele frequencies by adding a coefficient to adjust the relative survival or reproduction of each allele. Random changes in allele frequencies become more pronounced in smaller populations. Within small populations, rare alleles are more likely to be lost, and one allele is likely to become "fixed" (universal), so random genetic variation among individuals is reduced. On the other hand, random genetic variation among small populations is enhanced.

Concurrently with the theoretical advances, experimental studies in laboratories and quantitative studies of populations in natural conditions confirmed all of these processes (Dobzhansky 1937). It was found that individuals in a population often differ in survival or reproductive success, these differences are often heritable, and genetic variation depends on the sizes of populations. Hoekstra et al. (2001) and Kingsolver et al. (2001) provide reviews of the prevalence of selection in natural populations.

The theoretical study of natural selection and evolution has in recent decades developed great sophistication in exploring the manifold complexities of population size and structure, mating systems, social interactions, migration, and isolation. Empirical studies continue to document the relevant processes. These studies, especially in natural conditions, face challenges in verifying small effects of selection and complex contingencies, in conjunction with randomness in finite populations. These effects are just the sort that theoretical investigations tend to explore.

Despite these theoretical and empirical advances, natural selection still has its perplexities and confusions. Natural selection, those who study it agree, results from the correlated consequences of individual variation, heredity, reproduction, survival, and competition and produces adapted change in the composition of a population. In various contexts, these components have raised many contentious issues. Is natural selection the result or the cause of adaptations? What kinds of variation and heredity are affected by natural selection? How do survival and reproduction interact? What about cooperation as well as competition? Even more important, is natural selection fundamentally misleading? On the one hand, is it so simple that it reduces to a tautology and explains nothing? On the other hand, is there enough complexity to explain the emergence of cooperation, culture, language? Is it even specifically a biological process?

First, a clear definition is needed. Natural selection, along with mutation, migration, and drift (randomness), produces evolution. Evolution is a change in the genetic structure of a population of organisms. In the simplest case, it is a change in the frequencies of alleles in the population. Natural selection then occurs when individuals differ in their survival or reproduction in ways associated with differences in their alleles. It is important to point out that natural selection does not result merely from differences in survival or reproduction of individuals. It also requires heritability of those differences. Natural selection is thus a change in the frequencies of alleles in a population as a result of differences in the survival and reproduction of individuals that carry those alleles. It is a matter of arithmetic: in any population, genetic variants spread when they leave more copies in successive generations.

Such a definition resolves one basic issue above. Natural selection is not tautological. It is not survival of those that survive. It both results from adaptation (of individuals) and produces adaptation (of populations). The general principle is indeed simple and self-evident. If individuals with particular features survive and reproduce better than others (call these individuals adapted) and if reproduction preserves features of the original, with occasional variation, then a population will accumulate adapted individuals. This concatenation of simple arithmetical steps is unarguable.

Heritable Differences

The mechanisms of heredity and variation have become progressively clearer and their complexities better understood. Within 50 years after *The Origin of Species*, the particulate inheritance of discrete features of plants and animals had become the subject of rapidly expanding research. Fifty years later, a century after the Origin, the molecular structure of a gene had been discovered. Now, some 60 years later, many complexities of molecular genetics have been investigated, although challenges remain.

The basis for heredity is an organism's genome, strands of DNA of enormous length duplicated in each of its cells. Segments of this DNA encode the amino acid sequences for several tens of thousands of different proteins, which compose much of each cell's structure and regulate its vital functions. Other segments encode a large variety of RNA molecules, which themselves (without translation into proteins) provide a diverse array of regulatory actions. An organism's DNA also includes parasitic components, which hitchhike on the mechanisms for duplication or which subvert the mechanisms for translation for its own purposes. All of these direct and indirect effects of DNA provide mechanisms for heredity and opportunities for variation.

The genome, we now know, is not the only way that parents can transmit their features to their progeny (Jablonka and Lamb 2014; Robert 2009). The cytoplasmic contents of ova (and sperm, in special cases) are transferred to zygotes and influence their development. The DNA in mitochondria is the primary example of maternal cytoplasmic inheritance, but other components of the cytoplasm can also influence development.

Bonding of methyl groups to nucleotide bases in a segment of

DNA can decrease its rate of translation to proteins. This inactivation of DNA by methylation is catalyzed by enzymes encoded by DNA elsewhere in the genome. In some cases, methylation is also promoted by environmental conditions, such as temperature or stress. Furthermore, patterns of methylation are sometimes passed to progeny with the parents' DNA. Plants and animals differ in which nucleotide base is methylated (cysteine in animals, adenosine in plants) and also in the rates of transfer to progeny (greater in plants, which lack the isolated lines of germ cells in animals). Gradual loss of methylation in successive generations eventually attenuates its effects. Nevertheless, rates of methylation and rates of reversion vary markedly in different regions of DNA (Van der Graaf et al. 2015). Yet whether or not patterns of methylation persist across many generations, natural selection can enhance or diminish their influence on the activity of DNA, just as it can adjust other molecular mechanisms that regulate expression of DNA. Methylation, often called "epigenesis," meaning "beyond genetics," expands the possibilities for natural selection. With its susceptibility to environmental influences and its progressive loss, it provides a mechanism for heredity more flexible and less stable than other ways to regulate DNA.

Still less stable influences on development, but nevertheless hereditary, can result from direct responses to an environmental feature sustained across generations. Learned habits and customs are examples that can propagate in families and populations of interacting individuals. The seasonal territorial boundaries defended by many songbirds, as well as features of their songs, provide examples of learned information transferred across generations in organisms other than humans.

Consider even a suggestion by Lamarck that persistent abrasion of parts of the body could result in inheritance of calloused skin. This "inheritance" would occur in human populations, for instance, if children tended to follow parents' predominant activities, such as using hands for heavy work or bare feet for walking.

Direct environmental influences have often not been accepted as natural selection. Yet the capability for developing callouses, for instance, is likely to require a predisposition to respond to abrasion by thickening of the epidermis, and such a predisposition might depend on particular structural or regulatory proteins (or methylation of particular segments of DNA, or both), all of which would require particular variants of DNA. In other words, development of callouses would depend on a particular interaction of genes and environment. In this case, for a particular genetic structure of the organism, development of callouses would be especially sensitive to environmental conditions.

Other environmental influences on development, including learning, also require physiological mechanisms and predispositions to respond to features of the environment and thus are also subject to adaptation by natural selection.

The result can take different forms. Development might vary continuously with some environmental feature. Alternatively, developmental switches might produce several alternatives in response to particular environmental features. In other cases, development might be especially sensitive to an individual's genome, rather than to its environment. An example is human growth to a particular height, in a population of well-nourished individuals. In such cases, genetic influences are more or less "canalized" within a range of frequently encountered environments. Across the entire spectrum from predominant influences of the environment to predominant influences of the genome, the development of an organism is always an interaction of its particular genome and its particular environment.

The influences on an individual's development span a spectrum of stability from genome to environment. DNA, one of the most stable organic molecules known, retains some of its structure even in the remnants of organisms that died tens of thousands of years ago. In contrast, the most variable features of the environment, for example the weather, can hardly be predicted from day to day. The development of an individual depends on responses to this entire spectrum of influences. At one end is an archival plan, at the other an immediate context.

No successful construction can rely on one of these alone, neither plan nor context. Context alone has too many possibilities; a plan has too few. Trying to build without a plan is just as likely to fail as insistence on following a plan. Successful construction, as much as successful development, results from adaptations at various levels of stability and flexibility. A "tried and true" plan is important. So is attention to immediate context. Success requires stability across generations as well as flexibility in momentary responses.

Development of an organism, its construction, is thus a plan instantiated in a particular context. All contextual influences during the lifetime of an organism, whether temperature, nutritional, or sensory, are mediated by proteins encoded by the genome. An individual's response to any stimulus depends on its current state as much as it does on the impinging stimulus. An individual's behavior, for instance, is at any instant an interaction between its present state and the incident stimulation. This interaction of current state and immediate environment continues through the successive, incremental stages of development, throughout an organism's life.

Each individual's survival and reproduction thus result from an interaction of its current state and its current environment. This process occurring in all individuals of a population produces natural selection. So the progressive interaction of genes and environment during an individual's development is embedded in a longer interaction of genes and environment in the evolution of a population of individuals. The pattern of an organism's development is embedded in the pattern of a population's evolution. Natural selection, an interaction between genomes and environment directing the evolution of a population, results from interactions of genome and environment directing the development of each organism.

As a result of such pervasive interaction, we can draw four general conclusions about structure and context: (1) nothing is determined by structure, (2) nothing is determined by context, (3) everything is influenced by structure, and (4) everything is influenced by context. In these four conclusions, the general terms, structure and context, summarize a variety of more specific alternatives: plan and reality, physiological state and sensation, genotype and environment. Each pair of alternatives, substituted for structure and context in the four statements, produces equally general conclusions.

Variation in Heritable Features

Natural selection, as just described, might produce stability or flexibility in development, to any degree between extremes of "canalization" and "plasticity." In any particular case, the result depends on the nature of both genetic variation and environmental variation.

Genetic variation in a population is produced by mutations in the genome, by genetic drift (random variation in reproduction or survival), and by migration to and from other populations. R. A. Fisher (1930) first emphasized the importance of variation in his "Fundamental Theorem of Natural Selection," which states that evolution is proportional to genetic variation, for any strength of natural selection. G. R. Price generalized this equation, by partitioning the change in genotypes in a subsequent generation into the covariance between genes and environment. Price's equation makes it clear that this principle can apply to any change, including learning as well as evolution (Okasha 2008; Grafen 2015; Queller 2017). Variation is fundamental to natural selection just as it is to learning.

One consequence is that natural selection must depend on mutation rate. This rate determines the rate of increase in genetic variation in a population; the rate of decrease in genetic variation, in contrast, depends on random loss (genetic drift) and thus on the size of a population. Migration affects genetic variation also, but the principles are similar. Mutation rates, by their effects on natural selection, influence the rate at which a population adapts to environmental change.

The mutation rate at any locus in the genome depends on the regulation of duplication and repair of DNA by proteins encoded elsewhere in the genome. If so, mutation rates might evolve to adjust the stability of DNA at particular locations in the genome. Natural selection might adjust these rates to the rates at which relevant environmental features change. It is known that segments of DNA (and thus the corresponding proteins) differ in their mutation rates. Not well understood, however, is whether or not mutation rates themselves evolve to adjust the rate of evolution by natural selection at different places in the genome.

Environmental variation comes in a spectrum of periodicities, with durations from seconds to many centuries. The stability or plasticity of development or evolution depends on how natural selection responds to different degrees of environmental periodicity. Environmental variation with periods much shorter than an individual's life is best accommodated by direct influences of the environment on an individual's development. Environmental variation over periods of one or a few generations is often better accommodated by a few alternative sub-plans for development. Variation over intervals of many generations is handled most efficiently by revisions of the basic plan. For biological organisms, these three alternatives correspond respectively to learning, developmental switches (Pfennig 1990), and genomic encoding. These three alternatives are of course points in a continuous spectrum from flexibility to stability. Each of these developmental alternatives results from an interaction of environment and genome, with progressively decreasing reliance on environmental flexibility and increasing reliance on genomic stability.

Nevertheless, the entire spectrum of developmental alternatives rests ultimately on the genome, the most stable form of inheritance. The genomes of organisms must encode the capabilities and predispositions for genetic or environmental stability, for developmental switches, for temporary methylation, or for flexible learning – for development in response to long-term, medium-term, or short-term variations in the environment. The development of individuals thus cannot be separated from the evolution of populations. Natural selection occurs at all periodicities of environmental variation.

Cooperation and Competition

The evolution of cooperation has created another challenge for natural selection. Darwin's initial summary of natural selection emphasized a "struggle for existence," inspired by Malthus' observation that reproduction can outrun resources for survival. This "struggle" implies competition. Ecologists now recognize two forms of competition, aggressive and exploitative. In the first case, direct interaction between two individuals results in greater access to a limiting resource for one of them. In the second, one individual acquires proportionately more of a resource as a result of its greater efficiency at locating or harvesting it, without any direct interaction with other individuals. Both forms of competition can result in the "struggle" Darwin imagined as the basis for natural selection. In his discussion of honeybees, Darwin acknowledged the challenge that such competition presents for the evolution of cooperation.

Simplistic explanations for the evolution of cooperation prevailed for more than a century after *The Origin of Species*. During the middle decades of the twentieth century, for instance, it was widely assumed that cooperation would prevail in a population whenever cooperating individuals gained an advantage over noncooperators. Furthermore, it was assumed that competition between groups of cooperators and groups of non-cooperators would lead to the evolution of cooperation.

A path-breaking book, *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, challenged these assumptions in their naive forms (Williams 1966, also Dawkins 1976). Cooperation might provide advantages to individuals and thus favor the spread of alleles associated with helping, but it is also necessary to consider possibilities for the spread of alleles associated with exploitation of cooperators. Such "cheaters" would have the benefit of accepting help from cooperators without the cost of reciprocating. Cooperators then face the prospect of becoming "suckers," by providing benefits to others that do not return them. Simple math shows that alleles associated with cheating spread at the expense of those associated with helping, even to elimination of the last helper alleles from a population. Group selection, the selection of cooperative groups in competition with noncooperators, thus appeared in a new perspective. Cooperation had to spread within groups before it could spread by competition between groups. Even if cooperation occurred mainly in small and relatively stable groups of individuals, so that some groups might by chance lack individuals with alleles for cheating and cooperators could prosper, these groups would remain vulnerable to any new alleles associated with cheating, which would then spread by natural selection to exclude cooperation.

Nevertheless, persistent cooperation has been documented for many kinds of organisms (Koenig and Dickinson 2016). It is clear, though, that special conditions apply. First, cooperation must spread within groups before it can spread by selection among groups. Second, reciprocity is a key to cooperation: costly helping must have compensating benefits in return. Third, alleles associated with cooperation can propagate within families as a result of kin selection, a special case of natural selection.

The evolution of cooperation under these constraints requires alleles associated with a behavioral tactic more complex than a simple heritable tendency to help. Perhaps the simplest effective tactic is to try helping occasionally but to continue only if reciprocation ensues ("win stay, lose shift" or WSLS). For instance, sedentary individuals, restricted to interacting repeatedly with a few neighbors, can evolve neighborhoods of cooperation, provided neighbors have alleles predisposing them to initial cooperation rather than cheating. More effective is a capability for identifying and tracking individual partners (Wiley 2012, 2015). Alleles associated with tit for tat (WSLS directed at recognizable partners) then permit a cooperative resolution of the prisoner's dilemma. With further behavioral elaboration, individuals might have capabilities for tracking multiple partners. Then alleles associated with such tracking (combined with those for predictable reputations of helping others) can make diffuse reciprocity advantageous (Nowak 2006; Nowak and Highfield 2011).

A different sort of advantage accrues to helping when it is directed toward genealogical relatives. In kin selection, an allele associated with helping kin can spread provided the cost to the helping individual is less than the benefit to its relative, devalued by the probability that the relative carries a copy of the same allele (in excess of the probability in the population at random). The mathematical condition is $C < rB$, where r is the coefficient of genealogical relatedness (the relevant probability when an allele for helping is rare). Individuals do not have to recognize kin directly; they can reliably interact with kin simply as a result of, for instance, birth in the same nest.

Investigation of a wide variety of animals and human societies shows that helping is frequently directed to close kin. Nevertheless, clear exceptions occur. Furthermore, it is rare that the quantitative condition, $C < rB$, is met. The latter difficulty can be overcome by a combination of some reciprocity in addition to close kinship. Kinship and reciprocity should complement each other in the evolution of cooperation by natural selection.

A further complexity can favor the evolution of cooperation within groups: policing. If cooperators join forces to punish or exclude cheaters, the extra cost imposed on cheating can make it less competitive in relation to cooperation. In this case, the cost to cooperators of policing must not reduce the benefits of cooperation too much. Also, avoiding the costs of policing (while yet enjoying its diffuse benefits) becomes a second-order form of cheating. Finally, if policing results in the expulsion of cheaters from a group, so cheaters face the possibility of receiving no benefit whatsoever from membership in the group, then selection might favor a form of stealth-cheating by sophistication in evading detection. Alternatively, super-cheating might consist of complete disruption of a group in the expectation that strictly individual competition might provide greater advantages for a cheater than expulsion from a cooperative group.

This spectrum of possibilities for the evolution of cooperation by natural selection involves increasing behavioral complexity. Some of the options thus might apply only to humans. For instance, although cooperative interactions with kin are widespread among nonhuman animals, only a few nonhuman primates have enough complexity of individual recognition to support the formation of reputations (Wiley 2012). Evidence for policing by animals, even primates, is also sparse (Flack et al. 2005; Beisner and McCowan 2013). On the other hand, neither theory nor field work has yet plumbed the complexities of helping and cheating, either animal or human.

Constraints on Natural Selection

In the early years of population genetics, a controversy arose between two of the pioneers in this field. Fisher's Fundamental Theorem suggested that natural selection would move populations in a particular environment toward ever greater adaptation provided a source of genetic variation, such as mutation, persisted. Sewall Wright argued, on the other hand, that natural selection usually moved populations toward a local optimum in an adaptive landscape with multiple optima. An adaptive landscape, as Wright imagined it, is a multidimensional map of the adaptation of organisms as a function of possible genotypes, in other words, of all possible combinations of alleles at every genetic locus (Wright 1932, 1986). An adaptive landscape in this sense applies to a particular environment. Only if each allele evolved independently would natural selection lead to a unique, maximally adapted genotype for this environment, as Fisher indicated.

Interactions among alleles, as Wright argued, make multiple local optima for genotypes nearly inevitable. Constraints on interactions of alleles at the same locus or at different ones would produce adaptive peaks in any environment. Optima in such an adaptive landscape result from trade-offs in the interactions of alleles at the same or different loci, in one or multiple individuals. Such interactions are frequent in genomes and populations. Any one protein often affects more than one cellular function or trait (pleiotropy), and any one trait or function is often influenced by more than one protein (epistasis).

A change in one trait might benefit survival or reproduction only if a concurrent change occurs in another trait. Furthermore, social interactions can involve traits with advantages for one individual but disadvantages for another, or traits with advantages only when present in both individuals concurrently.

Alleles associated with such traits often do not spread in a population when rare. For instance, during sexual selection, alleles associated with a female preference for a male trait do not spread unless their frequency in the population exceeds a threshold (or their genetic correlation with alleles for the male trait exceeds a threshold) (Lande 1981; Kirkpatrick 1982; Andersson 1994). In general, alleles associated with producing a signal cannot spread when alleles for responding to the signal are too infrequent, even if a response would benefit a signaler. Vice versa, alleles for responding cannot spread when alleles for signaling are rare, even if a response to a signal would benefit the receiver. Only in a population with enough of both sorts of alleles can they both spread (Wiley 2015).

The situation is even more constrained when rare traits have costs for individuals. For instance, when a preference takes time or a signal entails exposure to predators, these individuals are often subject to increased mortality when their counterparts are not quickly located. The associated alleles are lost from the population by natural selection even more rapidly than by random genetic drift alone. Furthermore, alleles for mate preferences interact in counter-intuitive ways with those for ecological differences (Servedio and Kopp 2012; Servedio and Bürger 2014).

Interactions between heritable variants, whether pleiotropic or epistatic, place constraints on evolution by natural selection. Thresholds and isolated adapted optima result. Only genetic drift or extraordinary mutations can move populations past hurdles or valleys where genotypes are associated with disadvantageous traits of organisms (phenotypes). Because alleles cannot persist unless organisms associated with them survive and reproduce disproportionately, thresholds and isolated optima are not easily surpassed. When multiple traits or genes are required for an overall advantage in survival and reproduction, the probability of overcoming disadvantages decreases.

In some cases, it is possible to circumvent hurdles by coopting unrelated traits or functions (Wiley 2017). In the course of natural selection, fins can become repurposed as legs and wings, just as ocelli can become photographic eyes, comfort movements can become signals, and perhaps habituation can become associative learning. Without cooptation of simpler traits to produce more complex ones, natural selection can overcome thresholds and isolated optima only by waiting for fortuitous mutations, either simultaneous combinations or discontinuous effects, occurrences sometimes called "hopeful monsters." Perhaps, in the long run, such events can occur. If so, evolution would inevitably lead to maximal adaptation of organisms, as Fisher indicated, despite the thresholds and local optima created by interacting alleles in the short run. Nevertheless, these interactions prolong, even when they do not prevent, evolution toward global optima by natural selection.

Empirical studies of natural selection have so far infrequently reported pleiotropy and epistasis of alleles under selection (Kingsolver and Diamond 2011). Perhaps these interactions in fact seldom constrain natural selection. On the other hand, the constraints might make selection more difficult to study, so that reports have focused on selection with little constraint. It is also possible that selection itself, over sufficient time, tends to reduce these constraints. For instance, duplication of a gene occurs frequently in the course of evolution, often followed by differentiation of the functions of the "daughters." Duplication and subsequent differentiation would reduce the constraints of pleiotropy on further progress of natural selection. The mechanism of duplication is itself regulated by other proteins and thus subject to evolution by natural selection.

Interactions between alleles at the same or different locations in DNA or between different consequences of the same allele, and the corresponding interactions between traits of organisms, all produce constraints on the progress of evolution by natural selection. As the complexity of organisms increases, it seems possible that these constraints become ever more complex and thus the constraints on evolution by natural selection ever more obstructive. Natural selection itself might produce still more complex genomes to reduce these constraints somewhat.

Evolutionary Computation

Biological evolution is not the only framework for discussing natural selection. Nothing precludes a generalization of its principles far beyond biological evolution. Optimizing structure by means of heritable variation and selection applies equally to evolution, epigenetics, and learning. In recent decades, it has also been applied to computation and molecular synthesis.

Evolutionary computing provides a way to optimize an algorithm (analogous to optimizing genetic structure) by systematically modifying its components (mutation) and then selecting those versions of components that optimize the output (organism or phenotype) for particular purposes (environments). The process is usually incremental and progressive like natural selection in biological evolution: mutation and selection occur repeatedly until a local optimum is reached.

An example is the use of neural networks to discriminate sets of inputs. In this case, a series of similar inputs is presented to a network of interacting nodes, each of which can promote or inhibit activity in other nodes and all of which combine to provide a response to each input. Randomly adjusting the interactions of nodes at each generation and then selecting those variants that improve discrimination between different sets of inputs eventually yields the best performance possible. In a similar way, pharmacologists search for optimal molecular structures by progressively altering the components of complex molecules (for instance, the sequences of amino acids in synthesized proteins) in order to maximize medical benefits and to minimize undesirable side effects.

Evolutionary computing or adaptive synthesis occurs in a multidimensional adaptive landscape just as biological evolution does. The adaptive landscape is the performance of an algorithm or synthetic molecule as a function of the hyperspace of possible structures (nodes and parameters or types and positions of chemical functional groups). Any solution encounters two widely discussed problems: too much precision to capture an entire adaptive peak; and too little accuracy to capture a global peak. In evolutionary computing, the first problem is called "over-fitting" (Domingos 2012; Srivastava et al. 2014); it applies to algorithms that perform well on initial data but poorly on similar, but previously unseen, data. Such an algorithm has evolved to a local peak but too narrowly. In other words, an algorithm with over-fitting has learned some of the noise (non-generalizable features) in the initial data as well as some of the signal (generalizable features). It has learned too much.

Ways to reduce over-fitting include early stopping (as soon as errors on previously unseen data increase too steeply) and limitations on structural and parametric complexity (by reducing the number of nodes and interactions and their weights). Constraining weights of parameters often reduces over-fitting by reducing the unpredictability of responses to similar but previously unseen data. All of these procedures rely on testing algorithms with unseen data. In terms of an adaptive landscape, they require testing performance on nearby parts of the landscape.

The second problem of evolutionary computing also has its analogue in biological evolution. Natural selection moves structures toward adaptation to a local optimum in the adaptive landscape and thus can miss a global optimum. Any algorithm, just as any population of organisms, evolves adaptation only to those inputs, variable or not, that it encounters. The only way to be sure of finding a global optimum is to test performance throughout the multidimensional adaptive landscape of possible structures. To assure finding this maximum, in the most general case, would require understanding the complete multidimensional structure and connections of the entire universe, down to the last quark. No advance in evolutionary computing, even as quantum computers increase the speed and breadth of learning or adapting, can guarantee discovery of a global maximum (Niu et al. 2019).

These problems in evolutionary computing have parallels in biological evolution. The generality and specialization of algorithms, we noted, can be probed by varying the environment. For natural populations, such probing of adaptations occurs when an environment varies in time. To acknowledge this variation, it has been suggested that a better metaphor is evolution in an "adaptive seascape." Adaptations of organisms are thus like well-fitted algorithms, both of which perform well over a local optimum with a spectrum of periodicities in input. Yet they do not necessarily perform well in similar environments not previously experienced. Organisms, including human engineers, instead settle for adaptation to a (not too) local optimum.

Humans making decisions with the assistance of evolutionary computing have learned to extend the principles of natural selection. Making decisions based on trial and error, whether by brains alone or by brains assisted by machines, results in optimal responses to each of numerous inputs. It is the basic process of human behavior. Indeed all animals, not just humans, learn to match responses to inputs. Decisions occur whenever an organism discriminates between alternative inputs when choosing what to eat or where to go or whom to associate with or to imitate. They do so because decisions in response to unpredictable inputs allow greater specificity in adaptations. An organism's capabilities and predispositions are specified by a stable plan, the organism's genome. Such a plan, as discussed above, is the basis for all forms of learning and culture. This plan then develops in conjunction with its immediate context, the organism's environment. Evolutionary computing is thus itself a result of evolution by natural selection.

Conclusion

Every organism develops from a particular plan in a particular environment. It persists as long as repair of its molecular components can counteract degradation – as long as its immediate structure can harvest exogenous energy to counteract entropy. Each organism, each instantiation of its plan, eventually decays. Yet, provided an organism transmits its original plan to nascent progeny, a similar organism in a similar environment can develop anew. Provided organisms transmit their plans to progeny with some appropriate level of variability, natural selection can yield a lineage of organisms that persists indefinitely in an environment of complex changes. Organisms with adaptations for learning can improve their survival in environments with short-term variation. These adaptations can extend even to learning the principles of natural selection. In the end, entropy, the ultimate noise in decisions, prevents learning with infallible foresight – and prevents immortality.

In a population of comparable entities, natural selection is no more than the spread of heritable variants that replicate at a higher rate than others. Natural selection is arithmetic applied to differences. The principles are the same in all cases. The mechanisms of heredity vary across a spectrum of stability, from the relative inflexibility of the genome to the increasing flexibility of developmental switches, epigenesis, and learning, even to quantum computing. Each mechanism is optimized for a pertinent environment by selection itself. Natural selection is potentially constrained by interactions within and between the entities in a population. It leads to greater complexity whenever it can produce more precise and accurate adaptations. The scope of evolution by natural selection thus includes the evolution of culture, cognition, and language. It thus leads to brains so large that they strain the limits of skeletal adaptations. The scope enlarges still further to include even those decisions assisted by machines.

Finally, consider several misconceptions about natural selection. Each misconstrues issues addressed above. All contradict evidence or logic.

The first misconception claims that culture is distinct from biology and thus not subject to natural selection. On the contrary, environment and genome interact in the development of all organisms, including humans. All features of an organism, including their predispositions and capabilities for learning, are influenced by their genetic structure, just as all features of an organism are also influenced by their environmental context.

Another misconception is that natural selection cannot accommodate Lamarkian evolution, in other words inheritance of environmental influences on individuals. We now know that such environmental influences can affect progeny, but natural selection produces and regulates the necessary mechanisms for these influences.

A third misconception is that natural selection, inasmuch as it is a selection, implies the existence of a selecting agent. Darwin was aware of this difficulty with the term "selection." Clearly rejecting any such agent, he nevertheless felt there was no succinct alternative for the term. Despite any limitations of language, there is no agent of selection.

Finally, it is sometimes claimed that selected individuals are morally superior. On the contrary, natural selection results from the arithmetic of survival and reproduction of genetic variants in limited populations. It has no more moral implications than any other example of arithmetic. Morality (ethics) instead applies to human attitudes toward the various consequences of natural selection. Because such behavioral dispositions are influenced by genes and by context, they are themselves influenced by natural selection.

Natural selection is not the child of morality; instead, morality is the child of natural selection. And not only morality but also philosophy. In the end, natural selection produces not only a philosophy of biology, but also *a biology of philosophy*.

References

Andersson, M. B. (1994). *Sexual Selection*. Princeton: Princeton University Press.

Beisner, B. A., & McCowan, B. (2013). Policing in nonhuman primates: Partial interventions serve a prosocial conflict management function in rhesus macaques. *PLoS One*, 8(10), e77369.

Bulmer, M. (2005). The theory of natural selection of Alfred Russel Wallace FRS. *Notes and Records of the Royal Society*, 59, 125–136.

Darwin, C. (1859). *On The Origin Of Species By Means Of Natural Selection*. London: Murray. Facsimile edition (2001), Cambridge, MA: Harvard University Press.

Darwin, C. (1909). *The Works of Charles Darwin, Volume 10: The Foundations of The Origin of Species: Two Essays Written In 1842 And 1844*. (F. Darwin, Ed.). Cambridge, UK: Cambridge University Press.

Darwin, C. (1975). *Charles Darwin's Natural Selection: Being the Second Part of His Big Species Book, 1856–1858*. (R. C. Stauffer, Ed.). Cambridge, UK: Cambridge University Press.

Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.

Dobzhansky, T. (1937). *Genetics and the Origin of Species*. New York: Columbia University Press.

Domingos, P. M. (2012). A few useful things to know about machine learning. Communications of the ACM, 55, 78–87.

Fisher, R. A. (1930). *The Genetical Theory Of Natural Selection*. Oxford: Clarendon Press.

Flack, J. C., De Waal, F. B., & Krakauer, D. C. (2005). Social structure, robustness, and policing cost in a cognitively sophisticated species. *American Naturalist*, 165(5), E126–E139.

Glick, T. F., & Kohn, D. (1996). *Darwin On Evolution: The Development Of The Theory Of Natural Selection*. Indianapolis: Hackett Publishing Company.

Grafen, A. (2015). Biological fitness and the fundamental theorem of natural selection. *American Naturalist*, 186(1), 1–14.

Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hoang, A., Hill, C. E., ... & Kingsolver, J. G. (2001). Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences*, 98, 9157–9160.

Jablonka, E., & Lamb, M. J. (2014). *Evolution In Four Dimensions:*

Genetic, Epigenetic, Behavioral, And Symbolic Variation In The History Of Life (2nd ed.). Cambridge, MA: MIT Press.

Kingsolver, J. G., & Diamond, S. E. (2011). Phenotypic selection in natural populations: What limits directional selection? *American Naturalist*, 177, 346–357.

Kingsolver, J. G., Hoekstra, H. E., Hoekstra, J. M., Berrigan, D., Vignieri, S. N., Hill, C. E., ... & Beerli, P. (2001). The strength of phenotypic selection in natural populations. *American Naturalist*, 157, 245–261.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. *Evolution*, 36, 1–12.

Koenig, W. D., & Dickinson, J. L. (2016). *Cooperative Breeding in Vertebrates: Studies of Ecology, Evolution, and Behavior*. Cambridge, UK: Cambridge University Press.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. *Proceedings of the National Academy of Sciences*, 78, 3721–3725.

Maynard Smith, J. (1998). *Evolutionary Genetics* (2nd ed.). Oxford: Oxford University Press.

Niu, M. Y., Boixo, S., Smelyanskiy, V. N., & Neven, H. (2019). Universal quantum control through deep reinforcement learning. *npj Quantum Information*, 5, 1–8.

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314, 1560–1563.

Nowak, M., & Highfield, R. (2011). *Supercooperators: Altruism, Evolution, and Why We Need Each Other To Succeed*. New York: Free Press.

Okasha, S. (2008). Fisher's fundamental theorem of natural selection – A philosophical analysis. *British Journal for the Philosophy of Science*, 59, 319–351.

Pfennig, D. W. (1990). The adaptive significance of an environmentally-cued developmental switch in an anuran tadpole. *Oecologia*, 85, 101–107.

Queller, D. C. (2017). Fundamental theorems of evolution. *American Naturalist*, 189, 345–353.

Robert, J. S. (2009). *Embryology, Epigenesis and Evolution: Taking Development Seriously*. Cambridge, UK: Cambridge University Press.

Servedio, M. R., & Bürger, R. (2014). The counterintuitive role of sexual selection in species maintenance and speciation. *Proceedings of the National Academy of Sciences*, 111(22), 8113–8118.

Servedio, M. R., & Kopp, M. (2012). Sexual selection and magic traits in speciation with gene flow. *Current Zoology*, 58, 510–516.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.

Van Der Graaf, A., Wardenaar, R., Neumann, D. A., Taudt, A., Shaw, R. G., Jansen, R. C., ... & Johannes, F. (2015). Rate, spectrum, and evolutionary dynamics of spontaneous epimutations. *Proceedings of the National Academy of Sciences*, 112, 6676–6681.

Wiley, R. H. (2012). Specificity and multiplicity in the recognition of individuals: Implications for the evolution of social behavior. *Biological Reviews*, 88, 179–195.

Wiley, R. H. (2015). *Noise Matters: The Evolution of Communication*. Cambridge, MA: Harvard University Press.

Wiley, R. H. (2017). How noise determines the evolution of communication. *Animal Behaviour*, 124, 307–313. [see Chapter 5]

Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton: Princeton University Press.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics*, 1, 355–366.

Wright, S. (1986). *Evolution: Selected Papers*. Chicago: University of Chicago Press

Section II

Introduction
Difficult Issues in the Evolution of Communication

Because communication occurs between living organisms (albeit often assisted by inanimate objects), and living organisms evolve, and all aspects of evolution are quantifiable, we require mathematics to understand the evolutionary adaptations of communication. The math uses elementary calculus, nothing the least complicated for any mathematician. The solutions required numerical integration and careful analysis to evaluate the apparent optima. Yet in the end the mathematical deductions produced intriguing graphical results and conclusions that are easily understood, that are indeed intuitive.

The first essay in this section presents the mathematical analysis of what happens when noise occurs in communication. It will emphasize a definition of noise as errors by a receiver. Errors are responses with disadvantages for the receiver. A receiver makes a decision every time it checks its sensory input. Does the input justify a response or not?

There are four mutually exclusive kinds of response to signals in noise, two correct (advantageous) and two erroneous (disadvantageous). The probabilities of the two kinds of error cannot be altered independently. It is impossible to decrease one kind of error without increasing the other. As a result any receiver faces a double-bind, and a receiver cannot evolve responses that reach perfection. An optimal trade-off is the only option.

Furthermore, advantages for a receiver in communication produce disadvantages for signalers. And vice versa. Receivers and signalers can only evolve to a joint optimum, at which each does the best it can provided the other does so as well. Different circumstances for communication, especially different features of noise, can alter this joint optimum or the receiver's trade-off. Nevertheless, communication cannot escape noise to reach perfection.

The second and third essays in this section examine two difficult issues in this approach to the evolution of communication. The second essay considers the practicalities and the currently available results of applying Signal Detection Theory to animal and human behavior.

The final essay examines what it means to claim that communication is a transfer of information between individual organisms, human or nonhuman. On the one hand, transfer of information seems obvious, at least between humans. Nevertheless, there is a long tradition that the cognition involved transcends mere behavior. It might involve symbols, not just signals. It might not be possible for nonhuman organisms, or at least not for more than a few. It might in the end not depend on the physiology of brains and other organs.

On the other hand, in recent decades, some students of animal behavior have argued that animals do not transmit information but instead use signals to manipulate another's behavior to their advantage. Manipulation implies that no information is transmitted. A signal only elicits a reflex by the receiver that is advantageous to the signaler. This view neglects the co-evolution of signalers and receivers. It also misconstrues information.

The issue of information in communication condenses to whether or not individuals can communicate their states of mind. The final chapter takes no final position on the equivalence of mind and behavior. It does conclude, however, that if mind is behavior then all communication transfers information about minds; if mind is not behavior, then no communication transfers such information.

Chapter 7

A Receiver–Signaler Equilibrium

Introduction

Questions about the evolution of communication have proliferated in recent decades, since Dawkins & Krebs (1978) emphasized that signalers and receivers often have conflicting interests. Since then theoretical, observational and experimental studies have dealt with questions such as, do signals communicate information?, what ensures honesty?, how do signaler and receiver converge on similar meanings of signals?, and how can communication evolve when signalers have no benefits in the absence of appropriate receivers and vice versa? During the same decades, investigations of mate choice and prey choice have also proliferated. These interactions consist primarily of communication, so some of the same questions arise in their study. In this welter of recent work on the evolution of communication, almost none has considered the consequences of noise.

Noise has featured more prominently in research on signal design, the properties of signals that minimize attenuation and degradation and maximize contrast with the background. The underlying objective of this research has been to explain how signals might evolve to increase the efficacy of communication in noise (Wiley & Richards, 1982; Endler, 1992; Brumm & Naguib, 2009). Theory has also addressed the consequences of noise for the evolution of communication. This work has identified the principal manifestation of noise — variable responses to a signal. It turns out, however, that statistical variance in responses does not change the equilibria of evolution (the evolutionarily stable states), so long as the mean response does not change (Grafen, 1990; Johnstone & Grafen, 1992).

In parallel with this work on the evolution of communication, the theory of signal detection in noise has developed in the past half century into a vast literature (Green & Swets, 1966; Macmillan & Creelman, 1991; Macmillan, 2002). Originally applied to procedures for evaluating the performance of receivers in psychophysical experiments, it now provides the rationale for analyzing responses in a wide range of psychological studies.

Signal detection theory has more recently been extended to the evolution of receivers (Wiley, 1994, 2006). This approach suggests that the performance of receivers should evolve in accordance with the payoffs for erroneous and correct responses. It becomes clear that parameters critical for evaluating a receiver's performance are rarely if ever measured in studies of communication. Yet even if the theory of signal detection can help to explain the behavior of receivers, it cannot provide a complete explanation for the evolution of communication, because the optimal behavior of receivers depends on the behavior of signalers. Signalers influence the relationship of signal to noise for receivers, by altering the intensity, attenuation, degradation, and contrast of signals. On the other hand, the behavior of receivers alters the probability of responses to signals. The question, thus, remains: How does the interaction of signaler and receiver evolve?

The purpose of this paper is to combine the theory of signal detection with a model of signal production to explore the co-evolution of signaler and receiver. Because the ramifications of this topic are so numerous, the analysis here focuses on a particular case — communication in mate choice. It assumes the prevalent situation in which males produce signals to induce females to mate with them, and females can respond to these signals. Males, the signalers, incur benefits and costs as a result of producing signals, and females, the receivers, incur benefits and costs as a result of their responses. Unlike previous treatments of this situation, the present approach assumes that females must make their decisions in the presence of noise. In other words, females sometimes make errors in their responses to signals.

In this approach a receiver's optimal threshold for response depends on the intensity of the signal in relation to noise, in other words, the signal/noise ratio or the exaggeration of the signal. Conversely, the signaler's optimal level of exaggeration depends on the receiver's criterion for response, in other words, its selectivity or choosiness or, in simple cases, its threshold for response. A search of these optima reveals a joint optimum, a Nash equilibrium, at which each party does the best it can, provided the other does the same. The location of this joint optimum depends on the payoffs for the receiver and the signaler and on the probabilities of signaling and paying attention, by the signaler and the receiver, respectively.

Under plausible conditions for mate choice, there is a joint optimum with a higher threshold for a receiver (greater choosiness) and a higher level of exaggeration for a signaler than in other examples of communication, such as warning calls in the presence of a predator. At the joint optimum, communication overall is honest, although in particular instances of communication receivers remain susceptible to deception by inappropriate signalers and signalers remain susceptible to exploitation by inappropriate receivers (such as eavesdroppers, predators, or parasites). The evolution of communication in noise, thus, reaches a joint optimum that falls short of perfection. The equilibrium is not a Pareto point, at which neither party can do better. Receivers sometimes make mistakes, and signalers are sometimes frustrated.

Preview of the Argument

Communication in noise differs in a fundamental way from communication without noise, because a receiver faces four possible outcomes every time it checks its input. These outcomes present inevitable trade-offs for a receiver in adjusting its threshold for response. A signaler also faces trade-offs, in this case between costs and benefits as the exaggeration of signals increases. Furthermore, a receiver's and signaler's performances are mutually interdependent. The utility of a receiver's threshold depends on the signaler's exaggeration (the level of the signal in relation to the level of noise), and the utility of a signaler's exaggeration depends on the receiver's threshold. Diminishing returns for both receiver and signaler suggest the possibility of a joint evolutionary equilibrium for a receiver's threshold and a signaler's exaggeration.

The present analysis combines previous expressions for the utility of a receiver's threshold (U_r) and the utility of a signaler's exaggeration (U_S) in order to explore the possibility of this joint equilibrium. Utilities for both parties are expressed as survival × fecundity, an approximate measure of the spread of genes associated with a phenotype. Thus, U_r and $\mathrm{U_S}$, as functions of both the receiver's threshold (t) and the signaler's exaggeration (e), represent the adaptive landscapes for each party, and the reciprocal partial derivatives of these utilities, $\partial U_r/\partial e$ and $\partial U_S/\partial t$, approximate the selection gradients for the receiver's threshold and the signaler's exaggeration.

With parameters for both the receiver's and the signaler's performances set to plausible values for many cases of mate choice, the resulting analysis shows that there exists a joint optimum for the receiver's threshold and the signaler's exaggeration. This optimum is a Nash equilibrium at which neither party can do better by a unilateral change in behavior. In some conditions, the equilibrium for communication in mate choice occurs at a higher threshold and higher exaggeration than the equilibrium for communication with warning signals.

In general, these results indicate that the normal situation for communication in noise is honesty with deception — honesty on average but with instances of disadvantageous outcomes for receivers or signalers. Furthermore, the relationship between honesty and costs is more complex than currently recognized.

Most important, the joint optimum for receiver and signaler indicates that communication in noise cannot escape the problems created by noise. Noise is an inevitable component of communication, and perfection in communication is not expected in natural conditions.

The Equations for Sender and Receiver

1. The signal detection paradigm

The feature of signal detection that makes a joint optimum of signaler and receiver possible is the inescapable trade-offs faced by a receiver in deciding whether or not to respond (Wiley & Richards, 1982; Wiley, 1994). The characteristic of noise is two distinct kinds of error by receivers, errors of commission and omission. Noise is not just an increase in variance of responses. On the contrary, it is impossible to minimize the two kinds of error simultaneously. Decreasing the probability of one increases the probability of the other.

This trade-off is apparent in a diagram of signal detection in noise (**Figures 1 and 2**). A signal in this case is any pattern of energy or matter that evokes a response more often than randomly but does not provide all of the power for the response. Because the receiver provides some, often most, of the power for the response, the receiver must decide when to respond. A receiver must, therefore, consist of three components: a sensory mechanism, a mechanism to associate activity in the sensor with a particular response, and a mechanism to amplify the response. A receiver's sensor has a mean level of activity (with a variance) in the absence of a signal. A signal provides enough power to raise this level of activity, so that during a signal the activity in the sensor has a higher mean level and (if the signal includes its own variation) a higher variance (for further discussion of these points, see Wiley, 1994, 2006, 2013a, b).

A receiver in this situation must adopt some criterion for a response. The simplest criterion is a threshold (Figure 1). If activity in the sensor exceeds the threshold, the receiver responds. Otherwise, it does not. *Note that the receiver only 'knows' two states of the world* — sensor-activity-above-threshold or not. It is reasonable to presume that receivers can evolve a threshold at any level of sensor activity. Wherever the threshold is located, a receiver faces four possible outcomes each time it checks the activity of its sensor (in other words, pays attention) and decides to respond or not (*b*). If a signal is present and activity in the sensor is above threshold, the receiver responds, an instance of a correct detection (D). If activity in the sensor at that moment is below the threshold, the receiver fails to respond, a missed detection (M). When a signal is not present, two corresponding possibilities arise, either a false alarm (response but no signal, F) or a correct rejection (no signal, no response, R). Provided the distribution of activity by the sensor in the presence of a signal overlaps the distribution in the absence of a signal, there are four possible outcomes every time a receiver checks its sensor (**Figures 1 and 2**).

Inspection of **Figure 1** shows that a receiver can reduce its probability of a missed detection by lowering its threshold, but it thereby increases its probability of a false alarm. Raising its threshold can decrease false alarms but inevitably increases missed detections. Whenever noise and signal cannot be completely separated by the receiver's sensor, the two kinds of error cannot be concurrently minimized.

This model incorporates the essential feature of signal detection, the inevitable trade-off faced by a receiver. There are several points that need emphasis. First, noise is pervasive in communication. It is likely that all communication in natural situations occurs in the presence of overlapping distributions of noise with and without a signal. This expectation is reinforced by a result of the present analysis, which indicates that the joint optimum for signaler and receiver is unlikely to result in perfect communication. Diminishing returns of the approach to perfection guarantee noisy communication.

Second, an error by a receiver, in any analysis of the evolution of communication, is a decision that does not increase as much as possible the spread of its genes. An approximate measure of the spread of genes is the expected number of individual's genes in the next generation (its survival × fecundity). If a correct detection of a signal increases the receiver's survival or fecundity, but a missed detection or false alarm decreases them, then the latter two decisions are errors by the receiver.

Third, a receiver's criterion for a response can vary in complexity. A criterion for response might be a simple threshold, or it might be sophisticated human cognition. A criterion can be a highly tuned filter for particular features of stimulation. The complexity or selectivity of a criterion does not, however, change the inevitability of noise nor the trade-off between false alarms and missed detections (for more discussion of these points see Wiley, 1994, 2006).

2. The receiver's optimal performance

The first step in understanding the evolution of communication in noise is to find the optimal location of the receiver's threshold. To do so, it is necessary to define the overall utility of any threshold in terms of the receiver's survival × fecundity, the expected number of an individual's genes passing to the next generation. If fecundity and survival vary with the location of the receiver's threshold, then this product is a measure of selection on the location of the threshold. Because the four possible outcomes whenever a receiver checks its sensor are an exhaustive classification of mutually exclusive alternatives, the expected utility of a particular threshold is the sum of the probabilities of each outcome and its payoff (with each payoff expressed as survival × fecundity). The receiver's expected utility is thus:

$$U_r = p_S(p_\mathrm{D}d_r + (1- p_\mathrm{D})m_r) + (1-p_S)(p_\mathrm{F} f_r + (1-\mathrm{p_F})r_r)$$

where p_S = probability of a signal in a (usually brief) interval of time, p_D = probability of a correct detection (D) provided a signal has occurred, $1 - p_\mathrm{D}$ = probability of a missed detection (M) provided a signal has occurred, p_F and $1 - p_\mathrm{F}$ are analogous probabilities for a false alarm (F) and a correct rejection (R), the two possible outcomes when a signal has not occurred. d_r, m_r, f_r, and r_r are the payoffs for the four outcomes, D, M, F and R (**Table 1**).

Table 1.
Parameters for the analysis of communication in noise

(with default values for communication in mate choice
when not otherwise specified in the text)

Properties of noise

Mean level of noise in the receiver's sensor = 0
Standard deviation of noise = 1.0

Receiver's parameters

U_r Receiver's overall utility
d_r Payoff for a correct detection (D) = 2.0
m_r Payoff for a missed detection (M) = 1.0
f_r Payoff for a false alarm (F) = 0.5
r_r Payoff for a correct rejection (R) = 1.0
p_{D} Probability of a correct detection
p_{M} Probability of a missed detection (= $1 - p_{\mathrm{D}}$)
p_{F} Probability of a false alarm
p_{R} Probability of a correct rejection (= $1 - p_{\mathrm{F}}$)
t Location of a receiver's threshold (activity in a sensor > 0)
p_s Probability of a signal in a unit of time = 0.5 (see also below)

Signaler's parameters

U_s Signaler's overall utility
b_s Benefit as a result of a correct detection by a receiver = 2.0
n_s Benefit when a receiver does not respond to a signal = 1.0
s_0 Proportionate change in survival when **no** signal is produced = 1.0
c_m Marginal change in survival when a signal **is** produced = −0.01
s_s Survival as a result of producing a signal (= $s_0 + c_m e$)
p_s Probability of producing a signal in any unit of time = 0.5
e Exaggeration (level or magnitude) of a signal > 0

Alteration of a signal during transmission

(not included in the current analysis)

Attenuation (relative reduction of signal or exaggeration) = 1.0
Degradation (relative increase in signal variance) = 1.0

A receiver must receive a net benefit on average as a result of participating in communication, otherwise selection would eliminate responding to the signal. Consequently, some of the four outcomes must provide a positive payoff. Normally a correct detection would have the highest payoff in comparison to a correct rejection. In contrast, the two kinds of error, false alarm and missed detection, would often have adverse consequences and, thus, low payoffs in comparison to a correct rejection.

A receiver must receive a net benefit on average as a result of participating in communication, otherwise selection would eliminate responding to the signal. Consequently, some of the four outcomes must provide a positive payoff. Normally a correct detection would have the highest payoff in comparison to a correct rejection. In contrast, the two kinds of error, false alarm and missed detection, would often have adverse consequences and, thus, low payoffs in comparison to a correct rejection.

The optimal threshold for a receiver is the one that maximizes its expected utility, U_r. A previous analysis of the receiver's operating characteristic (ROC) showed that, for a particular signal/noise ratio, the optimal threshold depends on the slope of the indifference curve tangential to the ROC (Wiley, 1994):

$$(1 - p_S)(r_r - f_r)/p_S(d_r - m_r).$$

The optimal threshold is **high** when this slope is high and, thus, p_S and $(d_r - m_r)/(r_r - f_r)$ are low, and the optimal threshold is **low** when these parameters are high. A high threshold is termed 'adaptive choosiness', because missed detections are relatively frequent (but false alarms are infrequent). A low threshold is 'adaptive gullability', because false alarms are frequent (but missed detections are not) (Wiley, 1994).

A more general analysis, presented below, calculates U_r as a function of both the receiver's threshold (t) and the level of the signal (its exaggeration, e) in relation to the noise in a receiver's sensor,

$U_r = f(t,e)$.

For this analysis, the level of activity in a receiver's sensor in the presence of noise is assumed to have a truncated normal probability density function (PDF) with mean = 0 and standard deviation = 1.0 (**Figure 3**). Thus, levels of activity in the sensor when a signal is present are scaled with respect to a the standard deviation of noise in the sensor (a level of 2.0 in the presence of a signal means that the difference between the mean levels of noise and of signal plus noise is twice as great as the standard deviation of noise alone). The analysis assumes that a signal does not increase the variance (as opposed to the mean) of the activity of the receiver's sensor. In other words, it assumes there is no additional variation introduced by the signaler, by transmission, or by transduction in the sensor. This assumption is discussed further below. For any level of signal plus noise, it is possible to find the level of the receiver's threshold that maximizes its expected utility by solving the partial differential equation,

$\partial U_r/\partial t = 0$, e constant,

and checking the second derivative or inspecting $U_r = f(t)$ for all relevant levels of signal plus noise, e. Note that for every level of activity in the receiver's sensor, e, the probabilities of the four outcomes require recalculation. As a consequence the equation above can only be solved with numerical methods. Mathematica 8.0.4 was used to find these solutions. A combination of procedures D, FindRoot and Max yields the same results as procedure FindMaximum.

3. The signaler's optimal exaggeration

The signaler can evoke a response from an appropriate receiver by producing a signal with enough power to affect activity of the receiver's sensor. It is plausible to assume a proportionality between the level of the signal produced by the signaler and the level of activity in the receiver's sensor. Although signals are normally affected by spherical spreading and attenuation during transmission, nevertheless the power arriving at a receiver at any distance remains proportional to the power at the source (despite the disproportionate decrease in power with distance). The level of the signal at the source is, therefore, called its exaggeration.

The production of a signal plausibly incurs a cost, in energy expended, risks taken, or opportunities lost. These costs are likely to be (and are here assumed to be) proportional to the level of a signal, at least within some range of signal level. For any analysis of the evolution of communication, the cost of a signal should be measured in units of survival × fecundity. Challenging, although feasible, this task remains an objective for the future.

The present analysis assumes that producing a signal reduces the signaler's survival in inverse proportion to the exaggeration of the signal (**Figure 4,** top):

$$s_S = s_0 + c_m e,$$

where s_0 = survival when no signaling occurs and c_m = the marginal cost of increased signaling (= 0). By setting $s_0 = 1$, the actual survival becomes a proportion of the maximal survival in the absence of signaling ($s_S = s_0 = 1.0$ and $c_m = 0$). Because costs must also rise with the rate of signaling, the signaler's marginal cost of signaling is multiplied by his probability of signaling in any small interval of time, p_S. Recall that the probability of a signal also affects the receiver's performance.

A signaler receives a benefit (b_S) when an appropriate receiver responds in a way that raises the signaler's survival × fecundity. For instance, in the case of mate choice, a female's response might promote mating with a male signaler and, thus, an increase in the signaler's expected fecundity. In the absence of producing a signal, a male presumably would have a lower probability of mating and, thus, lower expected fecundity. Setting the signaler's survival × fecundity in the absence of a response (n_S) = 1.0 makes b_S proportional to the signaler's survival ×

fecundity in the absence of communication.

Note that the signaler’s utility is not strictly proportional to the exaggeration of the signal. Instead it depends on the receiver’s threshold in relation to the level of the signal, which fixes the probabilities of the four outcomes for the receiver. The higher the receiver’s threshold, the lower the probability of a correct detection and, thus, the lower the probability of a response to the signal. The present approach, therefore, calculates the expected utility for a signaler as a function of the receiver’s threshold and the level of exaggeration of the signal:

$$U_S = p_S s_S(p_{\mathrm{D}D} b_S + (1 - p_{\mathrm{D}})n_S) + (1 - p_S)s_0 n_S$$

where $s_S = s_0 + c_m e$, as above, p_{D}, p_{F}, d_r, m_r, f_r and r_r are the probabilities and payoffs of the receiver’s outcomes, as described in the previous section, p_S is the probability of signaling in a small unit of time, b_S is the benefit received from a response by the receiver, and n_S is the benefit received when there is no response (Table 1). Notice that this formulation assumes that the signaler receives no benefit from a false alarm. In mate choice, a false alarm by a receiver would consist of mating with a partner other than signaler.

For any level of the receiver’s threshold there exists an optimal level of signaling (exaggeration) by the signaler, the level that maximizes the signaler’s expected utility. At lower levels of exaggeration, the signaler evokes too few responses, and at higher levels, it incurs too high a cost in survival. The optimal level of signaling (exaggeration) as a function of the receiver’s threshold can be calculated by finding the solution to the partial differential equation,

$$\partial \mathrm{U}_{\mathrm{S}}/\partial \mathrm{e} = 0, \mathrm{t} = \text{constant},$$

and checking the second derivative or inspecting the contour of $U_S = g(e)$ for constant t. This solution can only be found by numerical methods, again as implemented in Mathematica 8.0.4 (see above).

4. The receiver's and signaler's joint optimum

So far this extension of signal detection theory has derived the overall utilities for a signaler and an appropriate receiver. Each of these utilities is a unique function of both the receiver's threshold and the signaler's exaggeration:

$$U_r = f(t,e); \quad U_S = g(t,e).$$

To find any joint optimum, it is necessary to search for points at which the receiver's optimal threshold and the signaler's optimal exaggeration coincide. These joint optima occur at the intersections of the two curves, $t^* = f(e)$ and $e^* = f(t\ )$, with an asterisk indicating an optimum. A joint optimum represents a particular combination of signaler's exaggeration and receiver's threshold that produce local maxima for both parties' utilities. A joint optimum is, thus, a Nash equilibrium for an interaction with the relevant parameters. Each party would do less well by unilaterally perturbing its behavior.

Depending on the receiver's and signaler's parameters, there were 0–2 such joint optima, as explained below. In all cases with two joint optima, one had lower utility for both sender and receiver and occurred at a combination of lower exaggeration of the signal and lower threshold by the receiver. To find the unique optimum (in cases with just one) or the more advantageous optimum (in cases with two optima), the present implementation in Mathematica 8.0.4 searched the level of exaggeration downwards to find the point at which (1) the two parties' optima coincided within a precision of <1% and (2) they both had higher utility than any second joint optimum.

Once a procedure was available for finding the most advantageous joint optimum for signaler and receiver, it was possible to explore the sensitivity of this optimum to perturbations of the parameters. This analysis explores in particular the relative magnitudes of the receiver's four payoffs and the signaler's cost and benefit. In each case, plots of a series of joint optima show how the joint optimum changes as each parameter changes. Because the possibilities are large, the present analysis focuses on situations that seem plausible for many cases of mate choice. For comparison, there is also briefer consideration of plausible cases of a warning call in the presence of a predator.

5. Communication during mate choice and warning calls

The receiver's parameters are predicted to differ contrastingly in these two situations. As described earlier (Wiley, 1994), mate choice is likely to result in a high threshold for response, adaptive choosiness, because false alarms by a receiver (choice of a suboptimal mate) have lower payoffs than missed detections (failing to respond to an optimal mate). A false alarm could result in a major reduction in a female's reproductive success, while a missed detection would result in continued searching, with some loss of time and exposure to risks, but with only a minor reduction in a female's reproductive success. In mate choice, the female's task is discriminating between optimal and suboptimal potential mates. The presence of the latter are the predominant forms of noise for this case of communication.

In contrast, warning signals are predicted to be associated with a low threshold for response and low exaggeration of signals, adaptive gullability (Wiley, 1994). In this case, a missed detection (failing to respond to a warning) would expose the receiver to a predator, while a false alarm (taking cover in the absence of a predator) would result in some loss of time, for instance for feeding or interacting with potential mates. The payoffs for missed detections and false alarms, therefore, contrast with the situation in mate choice.

Although mate choice and warnings illustrate contrasting payoffs for receivers, other forms of communication have their own relationships among the payoffs for the four possible outcomes a receiver faces. **Figure 5** is a proposal for arranging plausible relationships of these payoffs in different situations. The payoff for a correct rejection (no response when no signal is present), r_r, is set to 1.0, so that the payoffs for the remaining three outcomes are scaled to the expected utility of this one. The utility of a correct rejection is presumably similar to the utility of life in the absence of communication (no signals, no responses). With this scaling, the relative payoffs for remaining outcomes, along with the probability of a signal, determine the receiver's utility of participating in communication, relative to the utility of life in the absence of communication.

The following sections consider conditions for mate choice in which the payoff for a correct detection, d_r, takes values of 1.5, 2 and 3, while the payoff for a false alarm, f_r, takes values of 0.1, 0.5 and 0.9 (all payoffs relative to the payoff for R, as just explained). The payoff for a missed detection, m_r, is set at 0.9 (a 10% reduction compared to r_r as a result of lost time and increased risk of further searching). In contrast, conditions for warning signals have payoffs for correct detections of 0.8, 1.0 and 1.5, and for missed detections of 0.1, 0.5 and 0.9. The payoff for a false alarm in this case is set at 0.95.

This approach makes no attempt to justify these values because none has ever been measured. It is unlikely that all cases of mate choice or of warning signals would have relative parameters matching these figures. Nevertheless, these parameters seem plausible for at least some cases of mate choice and warning signals.

The signaler's parameters also have a large influence on the nature of communication. The present analysis considers a range of costs for producing a signal and benefits received by a signaler if an appropriate receiver responds. For mate choice, the signaler's benefit from a response, b_s, takes values from 1.5 to 8. The marginal cost of producing a signal, c_m, takes values of −0.001, −0.01 and −0.05. For warning signals, the signaler's benefit is set at 1.5 and its marginal cost of exaggeration at −0.01. The receiver's payoff for a false alarm, f_r, in this case is set at 0.99. The payoff for a correct detection (avoiding contact with a predator), d_r, takes values of 0.8, 1.0 and 1.5, and the payoff for a missed detection (lost time), m_r, takes values of 0.1, 0.5 and 0.9.

The payoffs for a false alarm and for a missed detection, therefore, contrast with the case of mate choice.

Another potentially critical difference between these two situations is the probability (1/frequency) of a signal, p_S. In this analysis, this parameter is set at 0.5 for mate choice (**Table 1**) and takes values of 0.01 and 0.001 for warning signals.

Default values for other parameters are presented in Table 1. In the cases of mate choice and warning signals considered here, it is assumed that a signaler does not benefit from a false alarm by a receiver (although in some cases of communication this possibility could arise). The analyses here also assume that the appropriate receiver is paying attention all of the time and is within range of the signaler.

Note that these models for noisy communication address the consequences for each instance of communication (each time a receiver checks its sensor or a signaler produces a signal). Depending on what constitutes a signal, many forms of communication can consist of hundreds or thousands of such instances in the life of an individual. On the other hand, some signals might occur once in a lifetime (constructing a display court, for instance).

Solutions for optimal thresholds and exaggeration

The first sections below present the utilities and optima for a receiver's threshold and a signaler's exaggeration for parameters that seem plausible for mate choice (Table 1). Then a comparison is made with plausible situations for warning signals.

1. Mate choice: the receiver's utility

For any set of parameters for the payoffs of the four possible outcomes for a receiver and for the probability of a signal, the receiver's utility is a function of its threshold for a response, t, and the signal level in relation to the noise, also called the exaggeration of the signal, e. U_r as a function of t and e is the adaptive landscape for a receiver's performance (Figure 6).

For any mean level of the signal (exaggeration), the utility of the receiver's threshold depends on its location (Figure 7). For a threshold close to 0 (the mean level of noise), the receiver's utility is usually low (≈1). A low threshold does a poor job of separating signal and noise, so many false alarms result. As the threshold increases, the receiver's utility increases to a maximum at some value below the level (exaggeration) of the signal. Higher thresholds result in a drop in the receiver's utility, because these thresholds exclude many correct detections.

Nevertheless, at high levels of the receiver's threshold, the receiver's utility changes only slightly with changes in the location of its threshold. The increased discrimination between signal and noise is offset by the decreased probability of correct detections. The drop is more pronounced the higher the payoff for a correct detection (**Figure 6**). It is also slightly more pronounced the higher the payoff for a false alarm, because then the cost of a mistake is less. Recall that all payoffs in these analyses are scaled in relation to $r_r = 1.0$, so $f_r < 1$ and $d_r > 1$.

A striking feature of the receiver's utility are the large domains in which it changes little with either the location of the threshold or the mean exaggeration of the signal. In these domains the trade-off faced by the receiver each time it decides to respond or not dominates its utility. Small changes in threshold or exaggeration result in counteracting changes in the probabilities of correct detections and false alarms. When the threshold < signal exaggeration, p_D decreases with increasing t less rapidly than does p_F ($\partial p_D/\partial t$ is less negative than $\partial\, p_F/\partial t$). When the threshold > exaggeration, this relationship reverses. The receiver's utility, thus, increases slowly as the threshold approaches the mean level of the signal, drops near this level, and then continues to drop slowly beyond the mean level of the signal. Overall the surface of U_r is relatively flat on either side of a locus of points along a diagonal line with a slope approximately equal to 1. **Figure 7** shows the optimal threshold for three levels of exaggeration (mean level of signal), when payoffs $d_r = 2$ and $f_r = 0.5$ (see Table 1 for default values of other parameters for mate choice).

For some sets of parameters, $U_r = f(t, e)$ has only a weak maximum (Figure 6, right), but in other cases, especially with d_r and f_r both small, there is a clear diagonal locus of maxima with $t < e$ (**Figure 6**, left).

2. Mate choice: the signaler's utility

For the signaler, the parameters affecting its performance (the marginal cost of producing a signal, the payoffs when an appropriate receiver responds or does not, and the probability of producing a signal) define its expected utility as a function of the exaggeration of its signal and the receiver's threshold. For any exaggeration (mean level of the signal), the receiver's threshold determines the probability of a response to the signal. The signaler's utility, $U_s = g(t,e)$, like the receiver's utility, $U_r = f(t,e)$, is a diagonal locus of maxima with e approximately equal to t (**Figures 8** and **9**).

For any level of the receiver's threshold, increasing the mean level of the signal at first increases the probability of responses (correct detections) by the receiver (**Figure 9**). The increased probability of responses is, however, balanced by the increased cost of producing a signal with a higher mean level. Near the point $t = e$, the increase in the probability of a response is greatest. A maximum is reached a point where $e \approx t$. Further increases in e result in a slow decline in the signaler's utility, as progressively less increase in responses is outweighed by a steadily increasing cost. **Figure 9** shows the signaler's optimal exaggeration for three different levels of the receiver's threshold, when $d_r = 2$ and $f_r = 0.5$ (other parameters have default values for mate choice, **Table 1**).

Like the receiver's utility, the signaler's utility includes large domains in which changes in t or e result in relatively little change in utility, as a result of trade-offs between costs and benefits. The locus of maxima is again a diagonal line with a slope approximately equal to 1. Note that the signaler's utility and its maxima do not depend on the payoffs for the four possible outcomes a receiver faces. It does depend on the probabilities of these outcomes, which are determined by the mean level of the signal (exaggeration) in relation to the noise.

3. Mate choice: the optima for receiver and signaler

Differentiating $U_r = f\,t, e)$ with respect to e and solving for $\partial U_r/\partial e = 0$ yields the locus of optimal thresholds for any set of parameters for the payoffs of the receiver's four possible outcomes and the probability of a signal (**Figure 10**, solid lines). These optimal thresholds either increase monotonically with exaggeration of the signal or in some cases have an abrupt concave shape as a result of a sharp rise in the optimal threshold for $e < 1$. As the figure shows, this concave shape arises when the payoffs for both false alarms and correct detections (f_r and d_r) are relatively low (**Figure 10**, top), so a high threshold at low levels of signal exaggeration avoids the high costs of false alarms. With a relatively high payoff (low cost) for false alarms, the optimal threshold remains 0 until signal exaggeration exceeds a minimal value near $e \approx (1, 2)$. Below this minimal exaggeration, it does not pay for the receiver to discriminate between signal and noise, because a threshold < 1 results in too many missed detections.

Note that when the optimal threshold = 0 it does not pay for a receiver to participate in communication. Instead, in these cases, it is better to respond regardless of the presence or absence of a signal. Above this minimal exaggeration, the receiver's optimal threshold increases approximately linearly with signal exaggeration with a slope < 1 and $t < e$, as previous inspection of the adaptive landscape for $U_r = f(t, e)$ suggested (**Figure 8**). At higher exaggeration of a signal, the receiver's optimal threshold diverges progressively from $t = e$. The lower tail of the PDF for the level of the signal always exceeds the upper tail of the PDF for the level of noise. In this region, $\partial t/\partial e < 1$ means that more of the signal is captured in relation to noise as exaggeration increases.

The signaler's optimal exaggeration for any set of its benefits and costs and the probability of a signal is obtained by differentiating $U_S = g(t, e)$ with respect to t and solving for $\partial U_S/\partial t$

= 0 (**Figure 10**, dashed lines). At very low thresholds for the receiver, it pays a signaler to increase the exaggeration of its signal rapidly. Above a value of the receiver's threshold near $t \approx (2, 3)$, the signaler's optimal exaggeration increases linearly with the receiver's threshold with a slope ≈ 1 but with $t < e$, as previous examination of the surface $U_S = g(t, e)$ suggested (**Figure 8**).

For any set of parameters, the joint optima for receiver and signaler occur where the lines of optima for each party intersect. By switching the axes for the signaler's optimal exaggeration, $e^* = f(t) \rightarrow t = f(e^*)$, and plotting the result with the receiver's optimal threshold $t^* = f(e)$, it is possible to visualize the joint optima where the lines cross at points $(t = t^*, e = e^*)$.

These plots (**Figure 10**) reveal three possible cases for these joint optima: either 0, 1, or 2 optima, depending on the parameters for the receiver's and signaler's performance. A single joint optimum occurs in those cases in which the locus of the receiver's optima is concave. As explained above, this case occurs when the payoffs for correct detections and false alarms, d_r and f_r, are both relatively low (low benefit for correct detection, high cost for false alarm).

Two joint optima occur with many sets of parameters, as a result of the upward curvature of the signaler's optima at thresholds near 0. One of the joint optima, thus, occurs with a low threshold and low exaggeration. The second joint optimum occurs at a much higher level of threshold and exaggeration, as a result of the steeper slope of the locus of optimal exaggeration. At each of these two points, neither party can improve its utility by perturbing its behavior (altering its threshold or the exaggeration of its signals, respectively). The lower point is often close to $t = 0$, so the receiver is close to no participation in communication at all (responding without regard to the presence or absence of a signal).

With some sets of parameters, the loci of optima for the receiver and the signaler do not intersect, and there is no joint optimum, although the lines of optima for the two parties converge and diverge as t or e increase. **Figure 10** suggests that this eventuality occurs when the receiver's payoffs for false alarms and correct detections are high (cost of a false alarm is low) and the signaler's benefit from a response is high (high d_r and f_r, low b_s).

The course of evolution through these joint adaptive landscapes as functions of t and e is best revealed by a plot of streamlines and vectors for the partial derivatives $\partial U_r/\partial e$ and $\partial U_s/\partial t$ (procedures VectorPlot and StreamPlot in Mathematica 8.0.4 produce **Figure 11**). The vectors in this plot (short arrows with the magnitude of the vector indicated by the size of the arrow) are the joint selection gradients on the behavior of receivers and signalers, as determined by the parameters of their performances (costs, benefits, probability of signals). The streamlines (long arrows that sum the vectors over longer trajectories) are, therefore, the expected trajectories of evolution. Notice that in all cases analyzed the arrows of evolution converge at a joint optimum corresponding to the upper optima in **Figure 10**. In some cases (**Figure 11**, upper right, lower left) at levels of t and $e < 2$, communication collapses as $t^* \rightarrow 0$. Notice that when the lines of optima do not cross (**Figure 10**, lower left, $b_s = 3, f_r = 0.9$), there is nevertheless an attraction point in the joint adaptive landscape (**Figure 11**, lower left), at a point above closest approach of the lines for the two parties' optima.

In all cases, the joint optima have asymmetrical slopes, with weak selective gradients on one side and strong ones on the other, a result of the large domains of nearly flat landscape for the functions, $U_r = f(t,e)$ and $U_s = g(e,t)$ (**Figures 6** and **8**). Nevertheless, from all directions around these joint optima, perturbations of either party's behavior would lower their utilities so that the selection gradients would tend to move their interaction back to the joint optimum. These points are, therefore, Nash equilibria for the interaction. A comparison of **Figures 6, 8** and **11** reveals that these equilibria are not necessarily Pareto optima, the points of maximal utility for either party alone.

Mate choice: influences of the receiver and signaler at the joint optimum

1. Receiver's payoffs

This initial analysis of noisy communication in mate choice explores the consequences of variation in four parameters: the receiver's payoffs for correct detections and false alarms and the signaler's marginal cost of signaling and benefit from a response by an appropriate receiver. The remaining parameters are set to default values (Table 1).

Recall that in communication for mate choice, a female receiver makes a correct detection when she responds to an optimal male and a false alarm when she responds to a suboptimal one. For simplicity the present analysis assumes only two categories of males, which produce two varieties of signals that are not always separated by females. The suboptimal males' signals, therefore, are the noise in communication between females and optimal males.

This analysis considered the following possibilities for the receiver's payoffs: d_r = {1.5, 2, 3} and p_F = {0.1, 0.5, 0.9}. As explained above, by setting the payoff for a correct rejection, r_r = 1.0, the remaining payoffs are scaled in relation to this one. By choosing an optimal mate, a female's payoff (her survival × fecundity) is 1.5, 2, or 3 times her payoff if she kept searching. By choosing a suboptimal mate, her payoff is 0.1 (high cost), 0.5, or 0.9 (low cost) times her payoff otherwise.

Inspection of the plots in **Figures 10** and **11** shows that, when a receiver's payoff for a correct detection increases, the upper joint optimum for communication moves to a lower threshold by the receiver and a lower exaggeration of the signal (summarized in **Figure 12**). Likewise, when a receiver's payoff for a false alarm increases (cost decreases), the joint optimum also moves to a lower threshold and a lower exaggeration. When the payoff for a correct detection is high and the cost of a false alarm is low, it pays for receivers to use low thresholds in order to increase the number of correct responses, despite a concomitant increase in the number of false alarms. Thus, the greatest exaggeration of signals and highest threshold for response occurs with a low payoff for correct detection and a high payoff for false alarm, in other words a situation in which the consequences of mate choice for a female are the least pronounced (little difference between the benefit of choosing an optimal mate and the cost of choosing a suboptimal one).

2. Signaler's benefit and cost

The analysis considered the following possibilities for the signaler's cost and benefit: c_m = {−0.001, −0.01, −0.05} and b_s = {2, 4, 6, 8}. The cost of producing a signal reduces the signaler's survival × fecundity by the marginal cost of exaggeration times the exaggeration of the signal, $c_m e$. The benefit a signaler receives when an appropriate receiver responds increases fecundity by the factor b_s. Recall that this analysis applies to each instance of communication, each time a receiver samples its sensor or a signaler produces a signal. Depending on what a signal is taken to represent, some of the marginal costs of signaling are unlikely. A signal, such as a bird's song, produced hundreds or thousands of times in an individual's life can hardly have a marginal cost of −0.05. On the other hand, constructing and maintaining a display court might represent a single signal with a marginal cost far greater than −0.002.

Inspection of the upper optima in **Figures 10** and **11** reveals that, not surprisingly, decreasing the marginal costs of signals or increasing the benefits of a response by a receiver increase both the optimal exaggeration of signals and the receiver's optimal threshold.

3. Comparison of mate choice and warning signals

The interest of this comparison, as explained above, comes from the contrasting relationship of the payoffs for the two possible errors by receivers, false alarms and missed detections. For warning signals, it is a missed detection that might have serious consequences, rather than a false alarm, as in the case of mate choice. In addition, the probability of a signal is often much lower for warning signals than for advertising signals. Finally, it is often difficult to identify the signaler's benefit from a response to a warning signal. In some cases, it might consist entirely of indirect benefits from kin selection. It is also possible that there is some direct benefit from notifying a predator that the signaler has spotted it.

For this analysis, it was assumed that a false alarm would cost little (have relatively high payoff) as a result of some time or opportunity lost for feeding or interacting with a mate (f_r = 0.95). The analysis then considered different payoffs for correct detections and missed detections by the receiver: d_r = {0.8, 1.0, 1.5} and m_r = {0.1, 0.4, 0.7}. A correct detection of a warning in the presence of a predator might have a payoff less than a correct rejection in the absence of a warning. Alternatively, it might have no effect or, if predators could strike without warning, it might increase survival. The consequence of a missed detection consists of exposure to a predator, so $\mathrm{m_r}$ = 0.1 would indicate dire consequences and 0.7 more modest ones. Also investigated were the benefit for the signaler provided the receiver responded, b_s = {1.2, 1.4, 1.8}, and levels of the marginal cost of exaggeration, c_m = {−0.001,−0.01,−0.05}.

Increases in d_r produced upper joint optima with lower exaggeration of the signal and lower thresholds for the receiver, just as in the case of mate choice above (**Figure 12**). Increases in m_r (higher payoff, lower cost) produced the opposite effect, upper joint optima with higher exaggeration and thresholds. Just as in the case of mate choice, a low payoff for a correct detection and a high one (low cost) for a missed detection resulted in the greatest exaggeration and highest thresholds. The plausible values for these parameters were less dispersed than for mate choice and, thus, resulted in smaller differences in the joint optima.

Again, not surprisingly, higher benefits for the signaler from responses by the receiver (b_s) and lower marginal costs of exaggeration (c_m) resulted in joint optima with higher exaggeration and higher thresholds. A lower probability of signals (p_s) also resulted in optima with higher exaggeration and higher thresholds. The highest exaggeration of signals for mate choice occurred with {low d_r, low f_r, low c_m, high b_s} and for warning signals with {low d_r, low m_r, low c_m, high b_s, low p_s}.

The actual values for optimal exaggeration and threshold in mate choice and warning signals were comparable in many cases. In both situations the highest joint optima were close to e^* = 6 and t^* = 3. These values, as explained earlier, are scaled to the standard deviation of noise in the receiver's sensor. Exaggeration = 6 is, thus, six times the standard deviation of the receiver's noise. The lowest upper optima have exaggeration and thresholds near e^* = 4 and t^* = 2. Two situations produced exceptionally high joint optima: mate choice when the marginal cost of exaggeration was low; and warning signals when the probability of a signal in the presence of a threat was low. Exaggeration of signals, according to this analysis, should be greatest under conditions that make measuring the marginal cost of signals most difficult.

Evaluation of the Results

1. Introduction

This analysis was intended to explore the consequences of noise for the evolution of communication. The inevitable trade-offs faced by both signalers and receivers during noisy communication frustrate simple intuitions. Do the trade-offs for receivers as well as signalers result in optimal thresholds for receivers and optimal exaggeration of signals? Do these individual optima ever coincide to produce joint optima for the interaction of signaler and receiver? Can noise explain differences in the exaggeration of signals in different circumstances? Can it explain the stability of honesty in communication? In the process of the investigation other issues arose as well. How are costs related to honesty in communication? What are the differences between mate choice and other forms of communication?

One result is clear. There is much more to learn about the evolution of communication in noise. The present model included only the minimal number of parameters to characterize signal detection in noise. Nevertheless, few of these parameters have ever been considered in studies of natural communication. The benefits to receivers of responding to signals have received some attention, but not the probability of correct detection. The costs and benefits to signalers have been addressed, but as discussed below it is clear that the potential complexities require much more investigation. Other parameters, the probabilities of the four possible outcomes for receivers, the payoffs for false alarms and missed detections, the probabilities of signals, have not been considered in studies of adaptations in communication.

It is just as surprising that these issues have never arisen in engineering applications either. There is a large body of work on optimal encoding of signals, but none that I know of on the costs and benefits of signal production and detection and their relationship. Yet the implications of noise for the evolution of communication apply just as well to the human design of communication. Most of the conclusions below apply to both evolutionary and economic scenarios.

As for the present model, although the number of parameters is minimal, it is nevertheless large. This report has only just begun exploring the consequences of variation in these parameters. The following sections address some of the questions raised above. They start with two old questions about the evolution of communication: the role of the signaler's costs and the stability of honesty.

2. Costs and benefits of signals and the stability of honesty

By formulating the costs and benefits of signals, the present analysis has clarified and also complicated previous conclusions about the role of costs in the evolution of honesty in signaling. If receivers cannot directly assess signalers' qualities and, therefore, only respond to the level (exaggeration) of their signals, and if signalers differ in intrinsic survival or marginal costs of exaggeration, signals can honestly indicate these aspects of quality. Figure 4 (top and middle) plots these relationships in a way that makes it clear that, if each signaler optimizes its level of exaggeration, by maximizing its survival × fecundity, then signals can honestly indicate each signaler's quality.

The present analysis adopted this approach for a signaler's cost as a function of the exaggeration of its signals. If intrinsic survival is the signaler's survival in the absence of signaling (s_0, survival when exaggeration = 0), then a constant marginal cost of exaggeration (c_m) results in a signaler's survival that decreases linearly with the exaggeration of signals:

$$s_s = s_0 + c_m e.$$

Although this graphical approach can clarify the relationship between a signaler's costs and honesty in signaling, it also raises some neglected questions. The absolute cost, relative cost, and marginal cost of signals differ in every case in **Figure 4** yet are rarely distinguished in discussions of the costs of signaling. In addition, there has been much discussion of different forms of 'handicaps', with an emphasis on whether or not costs of signaling are paid up front or not (Maynard Smith & Harper, 2003). This distinction can be captured by supposing that survival is a nonlinear function of signal exaggeration, concave either upward or downward. In addition, intrinsic survival and signaling costs might not vary independently.

Arguments about honesty also assume that the signaler's benefit is a monotonically increasing function of signal exaggeration. The analysis of noisy communication has shown how the probability of a receiver's response can increase monotonically with the signal's exaggeration, as a result of the receiver's adjusting its trade-off between missed detections and false alarms. Noise in communication is, thus, sufficient to explain a monotonic relationship between a signaler's benefit and the exaggeration of its signals.

Nevertheless, this relationship is not simple. A signaler benefits from a receiver's response, but the probability of a response depends on the location of the receiver's threshold as well as the exaggeration of the signal. The probability of a response, therefore, does not depend in any simple way on the exaggeration of the signal. For instance, as the level of a signal increases, the probability of correct detection falls more steeply than the probability of false alarm (see **Figure 1**). Thus, the proportionate change in correct responses, $p_D/(p_D + p_F)$, for any constant change in the level of a signal decreases with the level of the signal. In other words, a constant proportionate change in this ratio requires a larger proportionate change in signal level at higher signal levels, a result qualitatively similar to Weber's Law. More work is needed to examine the precise correspondence between the present model of a receiver's decisions in noise and other models of discrimination or decision (for instance, Kacelnik & Brito e Abreu, 1998).

Extending the graphical model of signaling suggests more complexity in the relationship between costs and honesty than previously supposed (**Figure 4**, bottom). For instance, honesty can result whether males differ in intrinsic (s_0) or in marginal survival (c_m) (see Getty, 1998; Wiley, 2000). A sufficient condition for this conclusion is that the functions for survival of signalers cannot cross. Yet this is not a necessary condition (**Figure 4**, bottom) when signalers with low intrinsic survival also have low marginal costs of exaggeration. This situation might arise if there were a developmental trade-off between intrinsic survival and exaggerated signals. How to interpret this situation would then depend on whether a signaler's quality was more accurately indicated by high intrinsic survival or low marginal costs of signaling.

Furthermore, males of different quality might accrue benefits at different rates. Suppose for instance, as a result of another developmental trade-off, that males with low intrinsic survival fertilize more eggs of females they attract. Their benefits of signaling would have a higher slope than that of males with high intrinsic survival. The interpretation of this situation would depend on whether quality was more accurately indicated by intrinsic survival or ability to fertilize eggs. The best indicator of quality, in every case, might instead be a male's expected survival × fecundity. So far as I know, none of these possibilities has received attention previously. Costs of signaling are related to honesty in communication in complex ways because of the interacting effects of a signaler's innate survival, marginal costs of exaggeration, and benefits of signaling.

An important conclusion from these suggestions and from the analysis of noisy communication is that the receiver's behavior is at least as important for explaining the exaggeration of signals and honesty in communication as are the signaler's costs. Indeed, the receiver's threshold, optimized in relation to the probabilities and payoffs of the four possible outcomes of any decision to respond or not, sets the conditions that determine the optimal exaggeration of signals and, thus, how much they cost.

The process of optimizing the individual parties' utilities during communication in noise often results in joint optima at which both parties benefit overall (with expected relative utilities > 1). These are Nash equilibria, combinations of behavior which neither receiver nor signaler can unilaterally perturb without decreasing its utility. They, therefore, represent stable conditions for communication with both parties benefiting. Nevertheless, at these equilibria receivers make errors, both false alarms and missed detections, and signalers do not always evoke responses from appropriate receivers. Such communication is, therefore, stable and honest on average, despite instances in which receiver or signaler or both do not benefit.

3. De novo *evolution of signals*

The plots of streamlines also clarify the selection gradients in the upper left corners, where the receiver's threshold is high and exaggeration approaches 0 (Figure 11). This is the situation for the initial evolution of a new signal. Presumably an incipient signal would have low exaggeration, and receivers would have little tendency to respond to it. Under these conditions, the selection gradients, although weak, uniformly stream to exaggeration = 0, or a collapse of communication.

The evolution of a new signal, therefore, has a hurdle to overcome. It requires either a preadaption or exaptation of a low threshold for response to the new signal, or it requires a behavior with an initial condition that already has a high contrast with noise. The first precondition might result from a sensory bias evolved in another behavioral context (Ryan & Keddy-Hector, 1992); the latter precondition might result from a previously irrelevant but conspicuous behavior, such as a displacement activity (Tinbergen, 1952). Either way it has long been recognized that the evolution of communication de novo must surmount a hurdle (for instance, Lande, 1981; Kirkpatrick, 1982, see Wiley, 2002, for further discussion).

4. Exaggeration of signals

This analysis of noisy communication has shown that the exaggeration of signals depends on the parameters affecting both the receiver's and the signaler's behavior. For both mate choice and warning signals different combinations of plausible parameters can produce at least 4-fold differences in exaggeration of signals and reach levels at least 6 times the standard deviation of noise in the receiver's sensor.

It is important to realize that this analysis did not include parameters for attenuation of a signal during transmission, nor for variation in the signal either at the source or as a consequence of propagation to the receiver. A parameter for attenuation would multiply the exaggeration of a signal at the source required to achieve any exaggeration at the receiver. The consequence of attenuation is, therefore, to raise the cost of signals for signalers, in order to achieve the optimal exaggeration at the receiver.

This analysis of noisy communication did not confirm the degree of contrast expected between levels of thresholds and exaggeration in signals in mate choice and warning signals (Wiley, 1994). Some cases of mate choice analyzed here do involve high levels of exaggeration and thresholds, but in many cases these levels broadly overlap those predicted for warning signals (Figure 12). This situation remains a conundrum for future analyses. If the relevant parameters were actually measured, would there be less contrast than we intuitively expect between such radically different situations for communication? Or are the parameters selected for this initial analysis in fact not realistic? Would other parameters produce greater contrast between mate choice and warning signals in noisy communication?

One result of including noise in any analysis of the evolution of communication is a clear prediction of the form that exaggeration should take. Relevant exaggeration consists of changes in a signal's properties that increase contrast with noise. Unlike other explanations for the spread of a receiver's responses to signals, noisy communication makes the prediction that receiver's responses favor the evolution of signals that contrast with noise from the position of the receiver. Investigation of adaptive signal design has suggested ways that signals can evolve to enhance contrast with noise (Wiley & Richards, 1982; Endler, 1993; Brumm & Naguib, 2009).

Theories of sexual selection do not usually include such predictions about which properties of signals should evolve. A few mathematical analyses have indicated that traits with greater advantages (or lesser disadvantages) for males should evolve preferentially (Heisler, 1984). Others have shown that traits promoting mate choice have evolved to increase contrast with the environmental background (Endler & Théry, 1996). An analysis of the evolution of communication in noise makes it clear that the exaggeration of preferred traits should usually follow this pattern.

The evolution of arbitrary preferences (responses to signals with no advantages for male or female) is the only exception to this rule. Such preferences cannot evolve if they have net costs (Pomiankowski, 1987). Any costs of searching must be balanced by benefits of mate choice. Considering the multiple payoffs and trade-offs for receivers in noisy communication, it seems unlikely that the effects of all the relevant parameters would exactly balance to yield no net cost nor gain ($U_r = 0$). Communication in noise, thus, makes it even less likely than otherwise that arbitrary responses and signals could evolve.

Although this analysis of noisy communication differs from sexual selection in predicting the direction of evolution for signals, it concurs with sexual selection in an important way. Despite the much fuller description of the interaction between a signaling male and a responding female, in the end this analysis shows how males with certain features mate with females with complementary features. This nonrandom mating generates a genetic correlation between any alleles associated with features of male signaling and those associated with features of female responding. This genetic correlation between signaler's and receiver's features can in certain circumstances generate accelerating (run-away) evolution of communication (Lande, 1981; Kirkpatrick, 1982). It is not clear whether or not Fisher had genetic correlation in mind when he described sexual selection. Nor is it clear whether or not he had frequency-dependent selection in mind, such as would apply to the evolution of all signals and corresponding responses, regardless of mating between signaler and receiver (Wiley, 2000). It is clear now that genetic correlation and accelerating evolution should apply to all cases of mate choice that meet certain initial conditions, regardless of whether or not signals are arbitrary or adaptive.

5. No perfection in communication

A fundamental conclusion from this analysis is that noisy communication is never perfect. Receivers and signalers instead evolve to joint optima, at which both parties benefit on average but at which both parties also fall short of perfection. Receivers remain susceptible to false alarms and, thus, to deception. Signalers remain incapable of evoking responses to every signal. Responding falls short of perfection and so does signaling. The system of communication is honest and stable, despite the occurrence of instances of communication disadvantageous to signaler or receiver.

From this point of view, the evolution of honesty in systems of communication is not surprising, but neither is the evolution of prevalent dishonesty and deception. This conclusion should apply to economic as well as evolutionary situations. We should not expect communication, of any sort, including systems designed by humans with costs and benefits in mind, ever to achieve perfection. The equilibrial condition for communication in noise is honesty with errors.

Noise is therefore an inevitable part of communication. By assuming that communication evolves in noise, this analysis shows that the evolution of communication cannot escape it. Evolution does not lead to signalers and receivers that perform ideally.

Noise is inescapable.

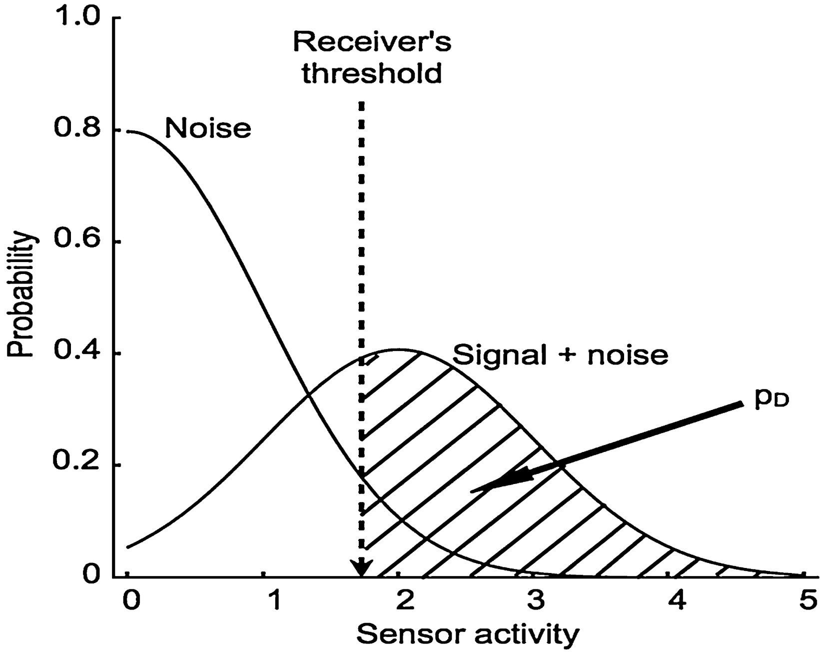

1.0
0.8
0.6
0.4
0.2
0
Probability
Receiver's threshold
Noise
Signal + noise
p_D
0
1
2
3
4
5
Sensor activity

Figure 1 (on opposite page). *An example of signal detection in noise*. The receiver's sensor has a probability distribution of activity for noise only and for noise plus a signal. In this example, these probability distribution functions (PDFs) have means = 0.0 and 2.0, respectively, and standard deviations = 1.0. The receiver sets a criterion for response, in this case a threshold level of activity in its sensor. This threshold in combination with the PDFs for noise and signal plus noise determine the probabilities of correct detection (p_D, the hatched area to the right of the threshold under the PDF for signal plus noise), false alarm (p_F, the area to the right of the threshold under the PDF for noise only), missed detection, and correct rejection (p_M and p_R, the areas to the left of the threshold under the PDFs for signal plus noise and for noise only, respectively). The hatched area, corresponding to p_D, provides an example of how one of these probabilities is calculated. If the receiver increases its threshold for response, p_F decreases but p_M increases (also p_D decreases and p_R increases). If it lowers its threshold, the consequences reverse.

		Receiver's decision	
		Response	No response
Signal	Present	CORRECT DETECTION	MISSED DETECTION
	Absent	FALSE ALARM	CORRECT REJECTION

Figure 2. *The exhaustive set of mutually exclusive outcomes* each time a receiver samples its sensor and decides to respond or not. A signal might be either **present** (stimulating the sensor in combination with noise) or **absent** (only noise stimulating the sensor). In either case, the receiver must decide to **respond** or **not**. If the receiver benefits on average from its decisions (usually because a correct detection has advantages for the receiver), then two of the outcomes (false alarm and missed detection) are usually errors with disadvantages for the receiver.

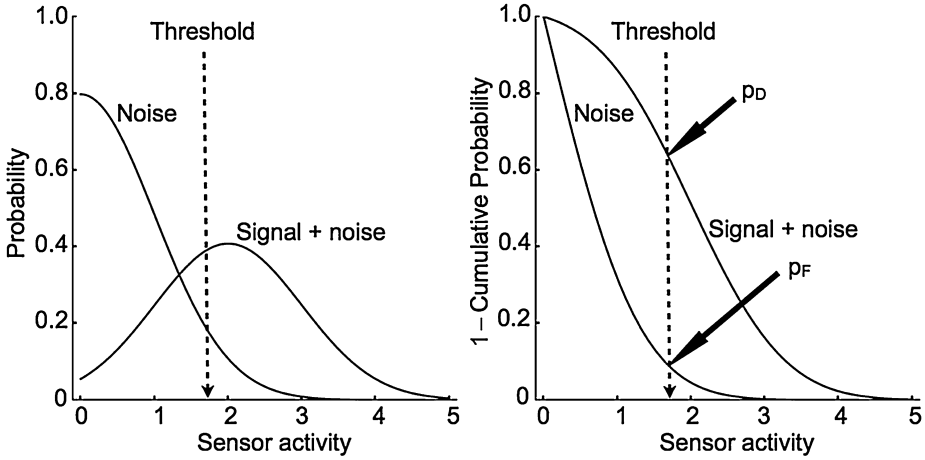


Figure 3. *Examples of truncated normal distributions.* The normal distribution (error function) expresses point probabilities for values between −∞ and +∞ and has a cumulative probability of all possible values = 1.0. Activity in any receiver's sensor, in contrast, only takes values >0. These truncated normal distributions express probabilities in proportion to the cumulative probability for values >0. This proportionality preserves the cumulative probability = 1.0 for all possible values of sensor activity. Possible probability density functions (PDFs, left) and cumulative density functions (CDFs, right) of activity in a receiver's sensor are shown for noise and for signal plus noise. An example of a receiver's threshold is also shown. The CDFs for a level of activity in the receiver's sensor at the threshold indicate p_F or p_D, in the cases of noise only or signal plus noise, respectively.

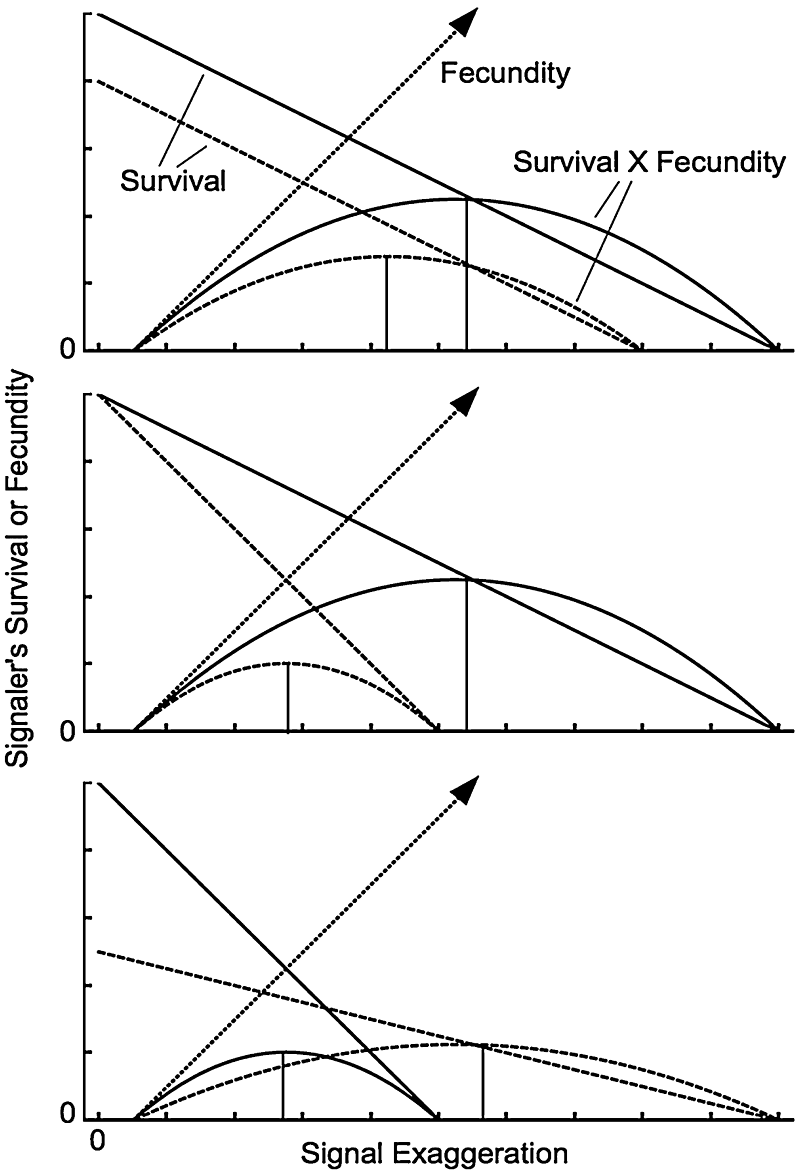

Fecundity
Survival
Survival X Fecundity
Signaler's Survival or Fecundity
0
0
0
0
Signal Exaggeration

Figure 4 (on opposite page). *Honesty in advertising* when signalers differ in intrinsic survival (survival in the absence of signaling, **top**) or in marginal costs of exaggeration (**middle**).

Each plot shows survival as a function of exaggeration for each of two signalers (sloping lines) and also their survival × fecundity (convex lines). Both signalers realize the same fecundity as a function of exaggeration of their signals, as would happen if receivers responded solely to signals and could not directly judge signalers' quality. **Vertical lines** indicate the level of exaggeration that would maximize each signaler's survival × fecundity. *Signalers of lower quality* (either intrinsic or marginal survival) *always have lower optimal levels of exaggeration*.

The scales of the axes are linear but otherwise unspecified; the vertical scale would usually differ for survival and fecundity (survival is always 1.0). Changes in scale do not affect the ranking of signalers' optimal levels of exaggeration.

The situation is more complicated (**bottom**) if the lines for signalers' survival cross or if signalers with lower intrinsic quality also have sufficiently lower marginal costs of exaggeration (see text for further discussion).

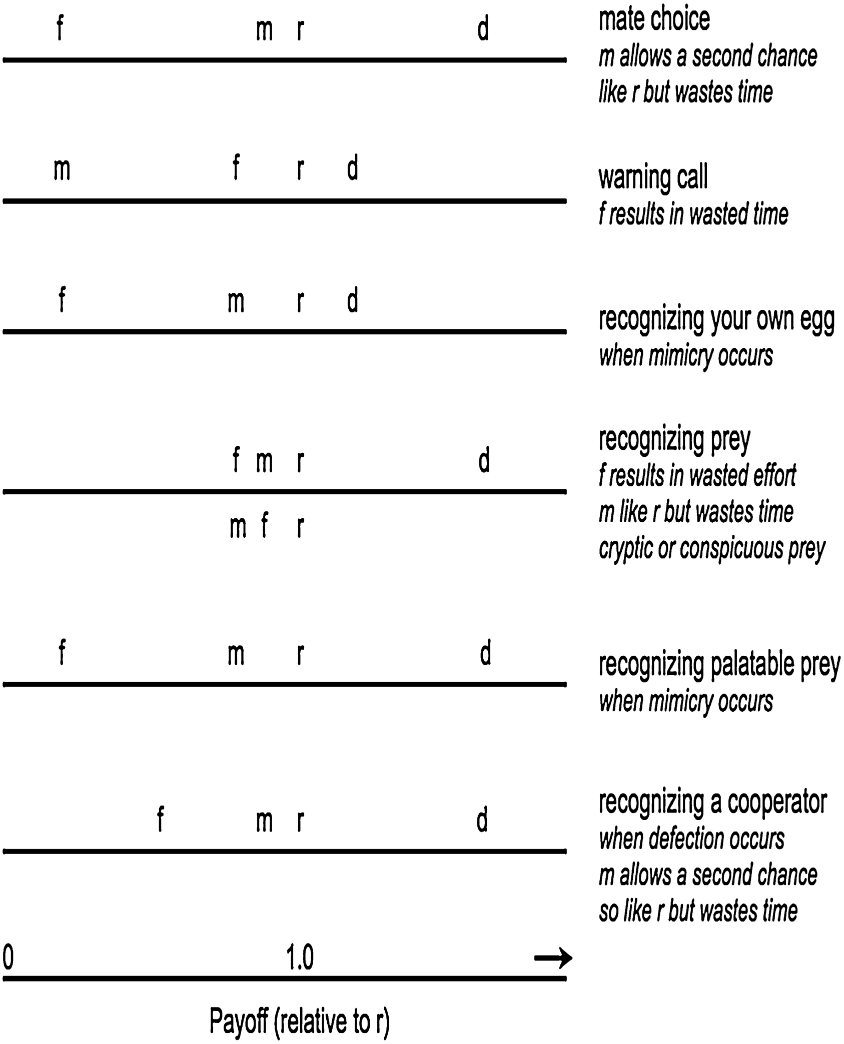

f m r d
mate choice
m allows a second chance
like r but wastes time
m f r d
warning call
f results in wasted time
f m r d
recognizing your own egg
when mimicry occurs
f m r d
m f r
recognizing prey
f results in wasted effort
m like r but wastes time
cryptic or conspicuous prey
f m r d
recognizing palatable prey
when mimicry occurs
f m r d
recognizing a cooperator
when defection occurs
m allows a second chance
so like r but wastes time
0
1.0
→
Payoff (relative to r)

Figure 5 (on opposite page). *Plausible values for the relative payoffs* for the receiver's four possible outcomes in a variety of situations for communication. The four outcomes (see **Figure 2**) are represented by d, m, f and r, and their relative payoffs in each situation are indicated by their positions on separate scales. Each payoff (advantage minus disadvantage for the receiver’s survival × fecundity) is proportional to the payoff for a correct rejection in the relevant situation. This payoff for no response when there is no signal is tantamount to life’s payoff in the absence of communication in this situation. With $p_{\mathrm{R}} = 1$, then usually $p_{\mathrm{D}} > 1$, $0 < p_{\mathrm{F}} < 1$, and $0 < p_{\mathrm{M}} < 1$. Within these limits, the magnitudes of plausible payoffs vary with each situation. Because only two points (0 and 1.0) are stipulated on these scales, the scales need not be linear. There are no measurements for all four payoffs in any case of communication that I know of, so the values suggested here are no more than plausible hypotheses.

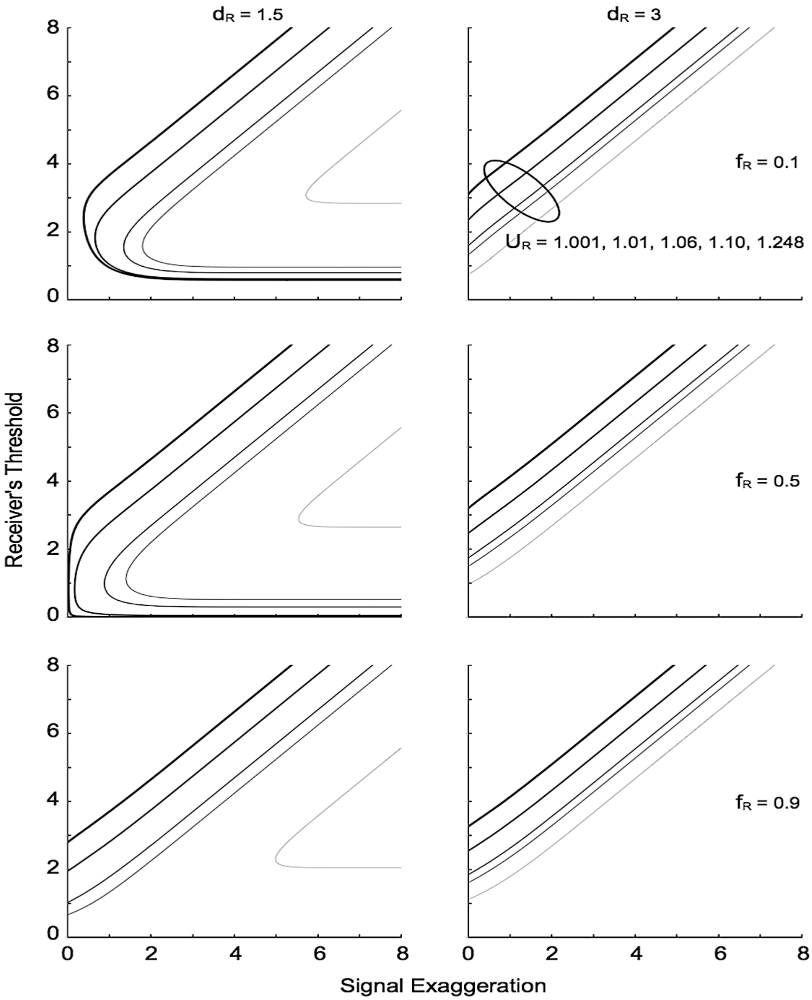


Figure 6. *Contours of the receiver's utility,* U_R, as a function of signal exaggeration and the receiver's threshold. The five contours represent (from thickest to thinnest) U_R = 1.001, 1.01, 1.06, 1.10 and 1.248, respectively. The highest value is close to the maximum for the conditions represented. The lowest value is set just above 1.0, because U_r > 1.0 over the entire plot in each case. The two columns show contours with d_R = 1.5 and 3.0; the three rows show them with f_R = 0.1, 0.5, and 0.9. (Continued opposite)

(Figure 6, continued from the opposite page) The contours show the relatively steep rise diagonally across each plot (with a slope ≈ 1) and in some cases also with very low thresholds. With d_R = 1.5 there is also an indication of a diagonal ridge of maxima. A similar ridge is much weaker with d_R = 3.0.

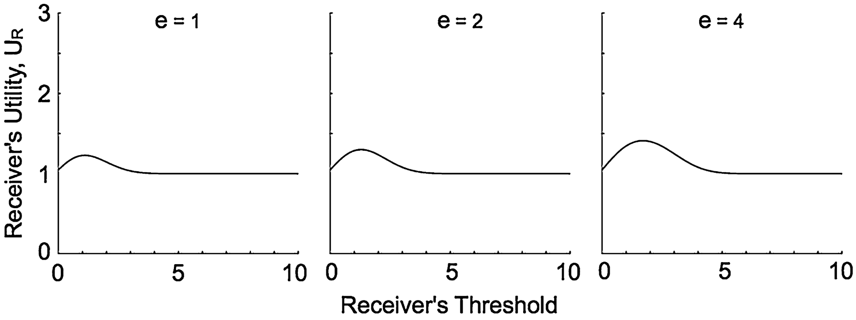


Figure 7. *The receiver's utility*, U_R, as a function of its threshold for three levels of signal exaggeration. For all plots, d_R = 2, f_R = 0.5, and other parameters have default values for mate choice (Table 1). In each of these cases there is a single maximum for the receiver's utility.

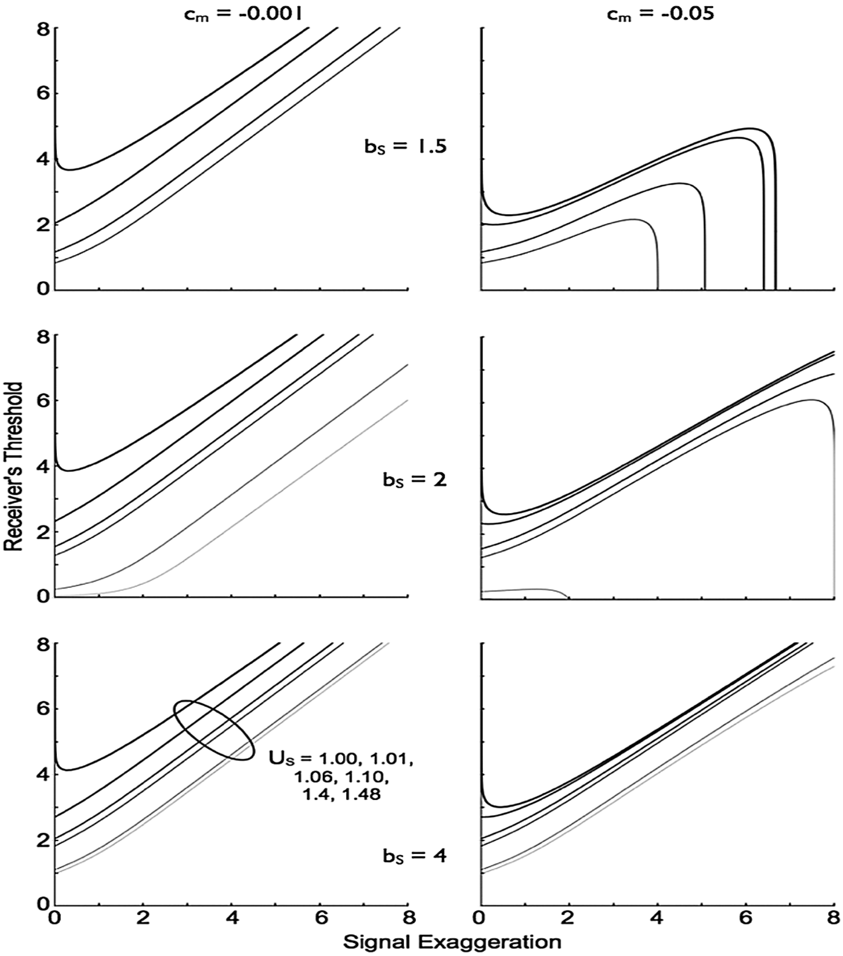


Figure 8. *Contours of the signaler's utility,* U_S, as a function of signal exaggeration and the receiver's threshold. The six contours represent (from thickest to thinnest) U_R = 1.00, 1.01, 1.06, 1.10, 1.4 and 1.48. The highest value is close to the maximum for the conditions represented. The lowest value at 1.0 indicates that in the upper left corner of each plot it does not pay for a potential signaler to produce signals (U_S < 1.0). The two columns show contours with b_S = 1.5 and 4.0; the three rows show them with c_m −0.001, −0.01 and −0.05. (Continued opposite)

(Figure 8, continued from the opposite page) The contours show the relatively steep rise diagonally across each plot (with a slope ≈ 1) and the limit beyond which further exaggeration does not pay.

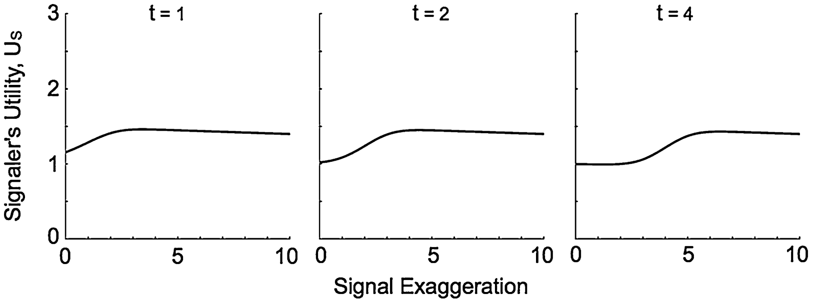


Figure 9. *The signaler's utility,* U_S, as a function of its exaggeration for three levels of the receiver's threshold. For all plots, $d_R = 2$, $f_R = 0.5$, and other parameters have default values for mate choice (Table 1). In each of these cases there is a single maximum for the receiver's utility, although the shoulders have low slopes. At high levels of the receiver's threshold, low levels of exaggeration do not pay ($U_S < 0$ with a slight negative slope).

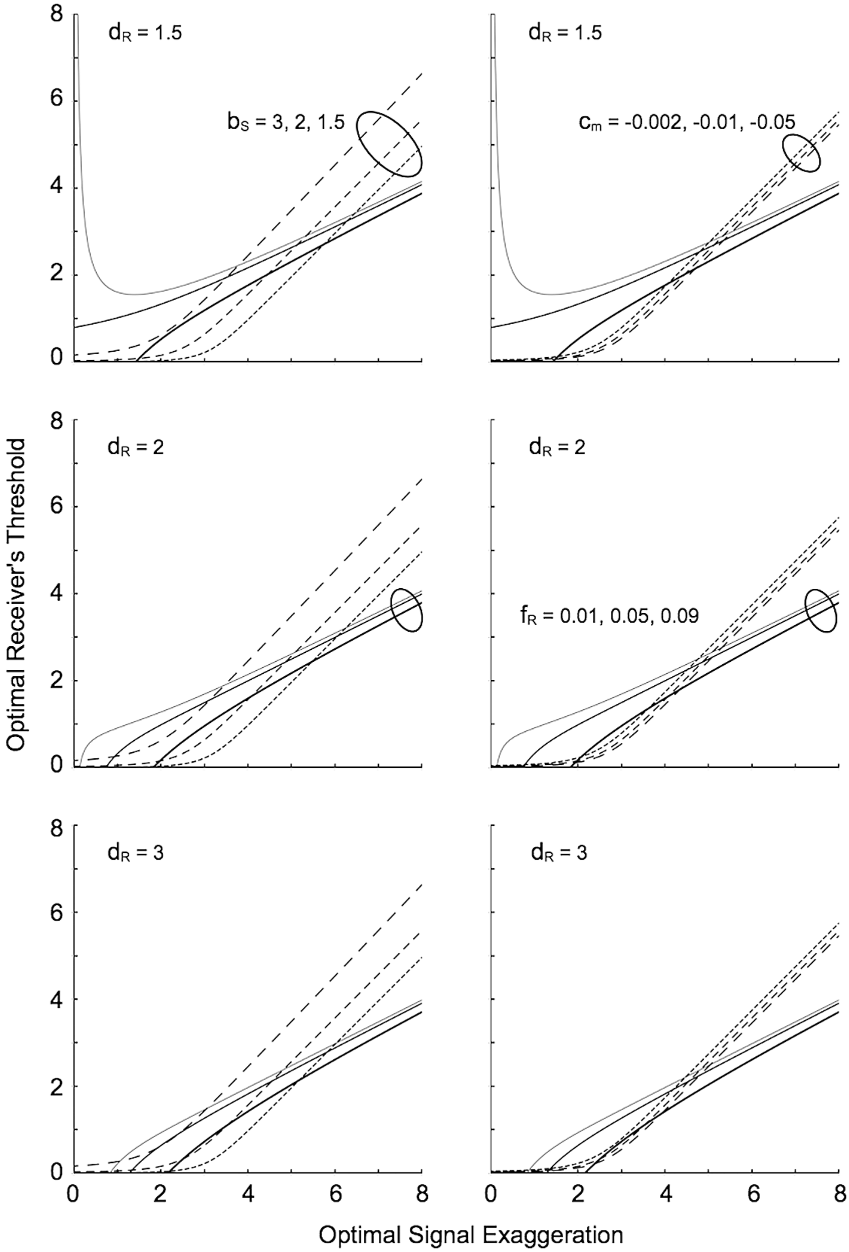

$d_R = 1.5$
$b_S = 3, 2, 1.5$
$d_R = 1.5$
$c_m = -0.002, -0.01, -0.05$
$d_R = 2$
$d_R = 2$
$f_R = 0.01, 0.05, 0.09$
$d_R = 3$
$d_R = 3$
Optimal Receiver's Threshold
Optimal Signal Exaggeration
0
2
4
6
8

Figure 10 (on opposite page). *The relationships of optimal thresholds***.** Each plot shows the locus of the signaler's optimal exaggeration as a function of a receiver's threshold (**solid lines**) and the converse, the locus of the receiver's optimal threshold as a function of the exaggeration of a signal (**dashed lines**).

In each plot the three solid lines show the receiver's optima for three values of the payoff for a false alarm (f_{R} = 0.01, 0.05 and 0.09 with thinner to thicker lines, respectively). The three rows show optima for three values of the payoff for a correct detection (d_{R} = 1.5, 2 and 3, respectively), the left column shows the signaler's optima for three values of the signaler's benefit from responses by the receiver (b_{S} = 3, 2 and 1.5 with thicker to thinner lines, respectively), and the right column shows the signaler's optima for three values of the marginal cost of exaggeration (c_m = −0.002, −0.01 and −0.05, respectively).

In many cases the loci for the receiver's optima and for the signaler's optima cross. These intersections indicate the Nash equilibria for a signaler and receiver under the respective conditions.

In some cases, the loci for optima do not cross, although they converge and diverge near points of attraction (see Figure 11). There are also intersections at very low values of thresholds and low values of exaggeration (see Figure 11 and discussion in the text).

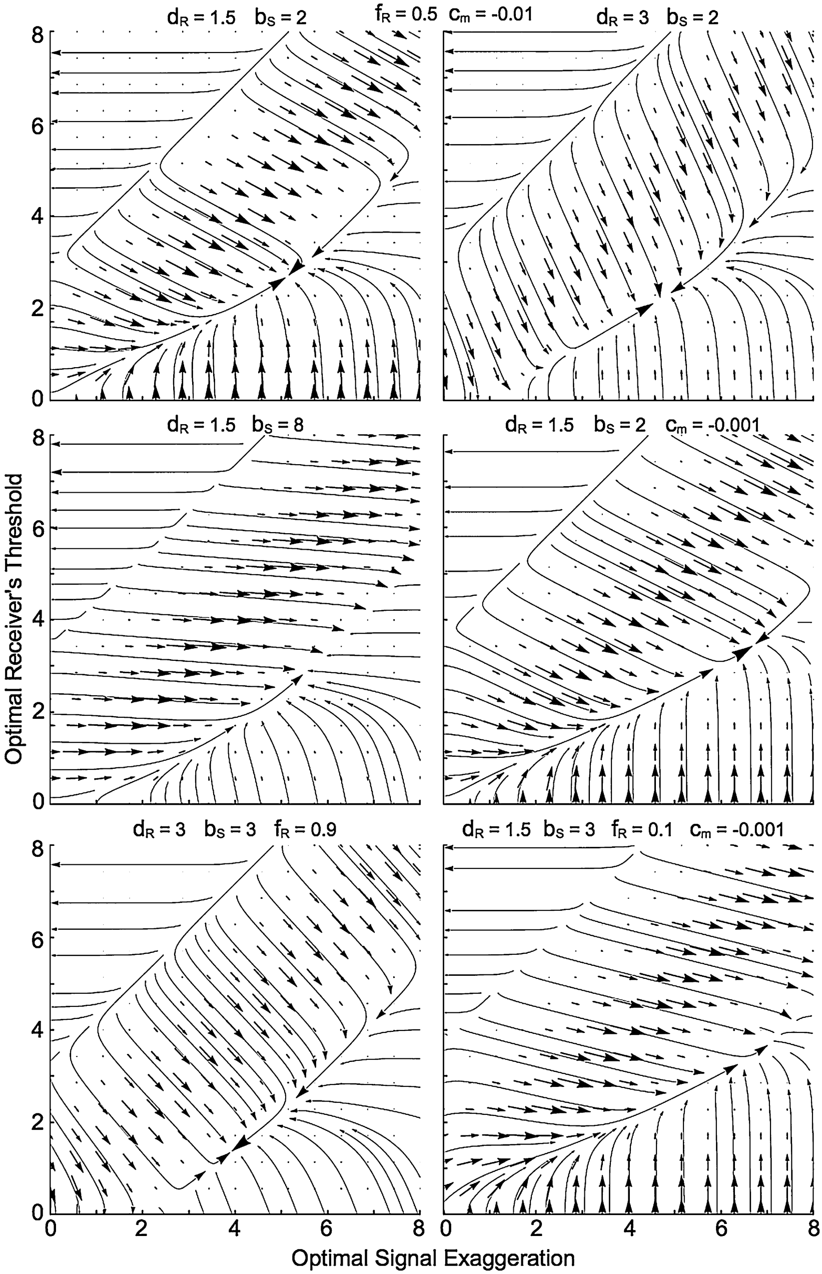

dR = 1.5 bS = 2
fR = 0.5 cm = -0.01
dR = 3 bS = 2
dR = 1.5 bS = 8
dR = 1.5 bS = 2 cm = -0.001
dR = 3 bS = 3 fR = 0.9
dR = 1.5 bS = 3 fR = 0.1 cm = -0.001
Optimal Receiver's Threshold
Optimal Signal Exaggeration
0
2
4
6
8

Figure 11 (on opposite page). *Streamlines and vectors* for the gradients of the signaler's and receiver's utilities as functions of the receiver's threshold and signal exaggeration.

The vectors (short arrows) show the gradients ($\partial U_r/\partial e$, $\partial U_s/\partial t$), with larger arrows for steeper gradients. The streamlines (longer arrows) result from sums of vectors. The vectors, thus, indicate the joint selection gradients on receivers and signalers under the conditions specified. The streamlines suggest the *trajectories of evolution*.

For all plots, the payoff for a false alarm, f_r = 0.5 and the marginal cost of exaggeration, c_m = −0.01. Other parameters are specified in the headings for each plot or have default values (**Table** 1). Points of convergence are joint optima for the receiver's threshold and signal exaggeration in each condition.

Compare these optima with the same ones displayed differently in **Figure 10**.

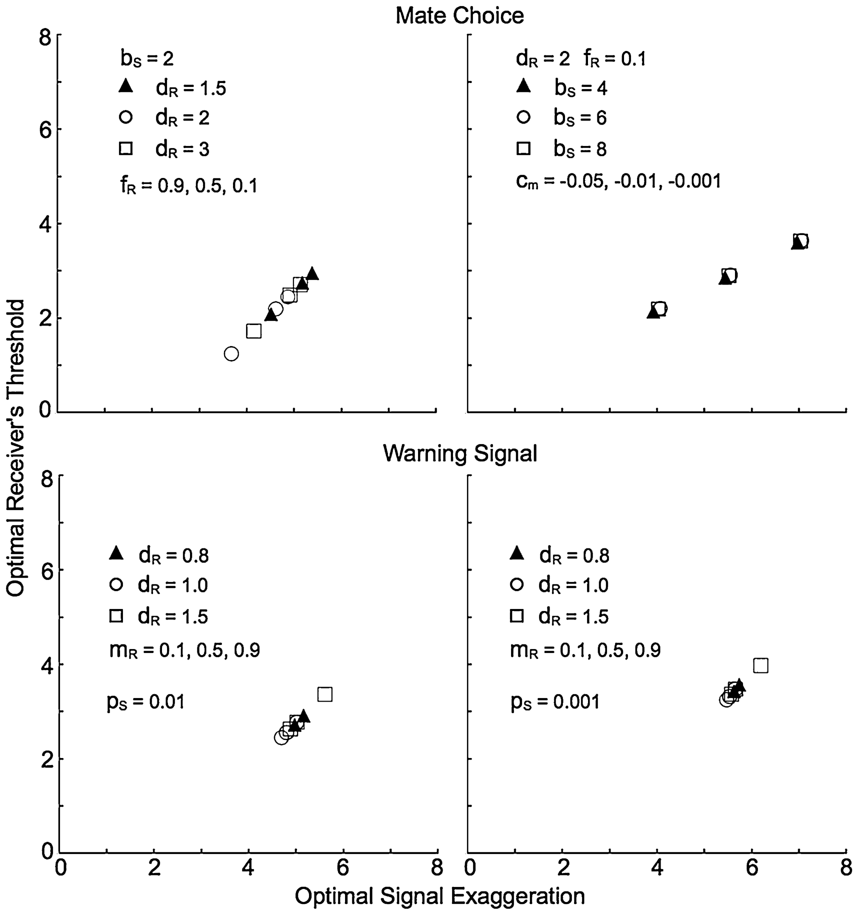

Mate Choice
Warning Signal
Optimal Receiver's Threshold
Optimal Signal Exaggeration
$b_S = 2$
▲ $d_R = 1.5$
○ $d_R = 2$
□ $d_R = 3$
$f_R = 0.9, 0.5, 0.1$
$d_R = 2$ $f_R = 0.1$
▲ $b_S = 4$
○ $b_S = 6$
□ $b_S = 8$
$c_m = -0.05, -0.01, -0.001$
▲ $d_R = 0.8$
○ $d_R = 1.0$
□ $d_R = 1.5$
$m_R = 0.1, 0.5, 0.9$
$p_S = 0.01$
▲ $d_R = 0.8$
○ $d_R = 1.0$
□ $d_R = 1.5$
$m_R = 0.1, 0.5, 0.9$
$p_S = 0.001$
0
2
4
6
8

Figure 12 (on opposite page). *The Nash equilibria for receivers and signalers.* The points show the upper equilibrial points in **Figure 10** for the receiver's threshold and the signaler's exaggeration for representative combinations of parameters for mate choice (upper two plots) and warning signals (lower two plots). Relevant parameters are specified in the legends; all other parameters have default values (**Table 1**).

References

Brumm, H. & Naguib, M. (2009). Environmental acoustics and the evolution of bird song. — *Advances in the Study of Behavior* 40: 1-33.

Dawkins, R. & Krebs, J.R. (1978). Animal signals: information or manipulation? — In: *Behavioural Ecology* (Krebs, J.R. & Davies, N.B., eds). Blackwell Scientific, Oxford, p. 282-309.

Endler, J.A. (1992). Signals, signal conditions, and the direction of evolution. — *American Naturalist* 139: S125-S153.

Endler, J.A. (1993). The color of light in forests and its implications. — *Ecological Monographs*. 63: 1-27.

Endler, J.A. & Théry, M. (1996). Interacting effects of lek placement, display behavior, ambient light, and color patterns in three neotropical forest-dwelling birds. — *American Naturalist* 148: 421-452.

Getty, T. (1998). Reliable signalling need not be a handicap. — *Animal Behaviour* 56: 253-255.

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

Green, D.M. & Swets, J.A. (1966). *Signal detection theory and psychophysics*. — Wiley, New York, NY (reprinted with additions by Krieger, New York, NY, 1974).

Heisler, I.L. (1984). A quantitative genetic model for the origin of mating preferences. — *Evolution* 38: 1283-1295.

Johnstone, R.A. & Grafen, A. (1992). Error-prone signalling. — *Proceedings of the Royal Society of London B* 248: 229-233.

Kacelnik, A. & Brito e Abreu, F. (1998). Risky choice and Weber's law. — *Journal of Theoretical Biology* 194: 289-298.

Kirkpatrick, M. (1982). Sexual selection and the evolution of female choice. — *Evolution* 36: 1-12.

Lande, R. (1981). Models of speciation by sexual selection on polygenic traits. — *Proceedings of the National Academy of Sciences USA*78: 3721-3725.

Macmillan, N.A. (2002). Signal detection theory. — In: *Stevens' Handbook of Experimental Psychology*, 3rd edition, Vol. 4 (Pashler, H.E., ed.). Wiley, New York, NY, pp. 43-90.

Macmillan, N.A. & Creelman, C.D. (1991). *Detection Theory: A User's Guide*. — Cambridge University Press, Cambridge (reprinted, Lawrence Erlbaum, Mahwah, NJ, 2004).

Maynard Smith, J. & Harper, D.G.C. (2003). *Animal Signals*. — Oxford University Press, Oxford.

Pomiankowski, A. (1987). The costs of choice in sexual selection. — *Journal of Theoretical Biology* 128: 195-218.

Ryan, M.J. & Keddy-Hector, A. (1992). Directional patterns of female mate choice and the role of sensory biases. — *American Naturalist* 139: S4-S35.

Tinbergen, N. (1952). “Derived” activities; their causation, biological significance, origin, and emancipation during evolution. — *Quarterly Review of Biology* 27: 1-32.

Wiley, R.H. (1994). Errors, exaggeration, and deception in animal communication. — In: *Behavioral Mechanisms in Evolutionary Ecology* (Real, L., ed.). University of Chicago press, Chicago, IL, p. 157-189.

Wiley, R.H. (2000). Sexual selection and mate choice: trade-offs for males and females. — In: *Vertebrate Mating Systems* (Apollonio, M., Festa-Bianchet, M. & Mainardi, D., eds). World Scientific Publishing, Singapore, p. 8-46.

Wiley, R.H. (2006). Signal detection and animal communication. — *Advances in the Study of Behavior* 36: 217-247. [see Chapter 8]

Wiley, R.H. (2013a). Communication as a transfer of information: measurement, mechanism and meaning. — In: *Information and Influence in Animal Communication* (Stegmann, U., ed.). Cambridge University Press, Cambridge, pp. 113-129. [see Chapter 9]

Wiley, R.H. (2013b). Signal detection, noise, and the evolution of communication. — In: *Animal Signals and Communication* (Brumm, H., ed.). Springer, Berlin, chapter 1.

Wiley, R.H. & Richards, D.G. (1982). Adaptations for acoustic communication in birds: sound transmission and signal detection. — In: *Acoustic Communication in Birds, Vol. 1, Communication and Behavior* (Kroodsma, D.H. & Miller, E.H., eds). Academic Press, New York, NY, p. 131-181.

Chapter 8

Signal Detection and Communication

Introduction

Although communication consists of associations between signals from one individual and responses by another, in reality these associations are often weak. In recent decades there has been a tendency to explain these weak associations as the result of attempts by signalers to manipulate or exploit receivers and of receivers to resist these attempts. This chapter takes a different approach, although the underlying question remains the same—how can signalers and receivers optimize their behavior? The present approach develops an earlier suggestion that it is the inevitability of errors by receivers that limits optimal behavior by both parties in communication (Wiley, 1994). Signal detection theory provides the basic theory for this approach. The previous applications of this theory, however, have been in psychophysics. To justify its application to the evolution of communication is the purpose of the present chapter.

The problems of signal detection arise especially for signals in their natural contexts. The properties of signals perceived by a receiver inevitably differ from those emitted by the signaler. For instance, acoustic signals like bird songs are altered by attenuation and degradation during propagation through the environment (Naguib, 2003; Naguib and Wiley, 2001; Wiley, 1991; Wiley and Richards, 1982). Although in any one situation, on average, some features of attenuation and degradation are predictable, much remains unpredictable in detail.

Furthermore, a receiver perceives this attenuated and degraded signal against a background of irrelevant energy that shares some features with the signal. An acoustic signal, for instance, is often perceived against a background of sounds with more or less similar frequencies, intervals, or other patterns. These sounds come from nearby individuals of the same or different species and from physical features of the environment such as wind and water. Finally, any receiver's sensory, associative, and motor neurons always include some unpredictability.

As a result of all of these processes, it is not surprising that signals usually have weak associations with responses. Sometimes when a stimulus occurs, the intended receiver fails to respond; sometimes the receiver responds when there is no stimulus. In the first case, the receiver seems to mistake a stimulus for the background; in the latter it seems to mistake the background for a stimulus. Because of the pervasiveness of these mistakes, receivers fail to achieve maximal performance and signals fail to reach maximal efficiency.

At first sight, these weak associations of signals and responses seem to be just noise in the system without fundamental implications for communication. This chapter, however, develops the view that these mistakes are a result of inescapable constraints on the performance of receivers and that these constraints in turn influence the evolution of both producing and responding to signals (Wiley, 1994). Many current issues in the study of communication, such as honesty and exploitation and the multiplicity and exaggeration of signals, become clearer once we understand the constraints on the performance of receivers.

These constraints on receivers are addressed by signal detection theory (Green and Swets, 1966; Macmillan and Creelman, 1991; McNicol, 1972). Developed originally by electronics engineers, in recent decades this theory has provided a rationale for the psychophysical study of sensory thresholds and perception. Despite its success in these studies, its application to the evolution of communication is still rudimentary.

The initial sections of this chapter provide an introduction to signal detection theory and its applications in psychophysics. The objective of these sections is to identify general principles for the study of adaptations in animal communication. These principles can clarify the properties of signals that affect a receiver's performance. They also suggest ways to extend the theory to the classification as well as detection of signals. These steps lead to hypotheses about the evolution of both signaling and receiving. In particular, signal detection theory leads to natural explanations for the evolution of deception and exaggeration in communication. The final sections take up the design and interpretation of experiments for studying communication in natural situations. The objective of these sections is to suggest practical ways to study the performance of receivers under conditions like those in which communication evolved.

Essential Features of Signal Detection

To apprehend the essential features of signal detection theory, it helps to consider a simple situation. Suppose an individual listens for a conspecific vocalization characterized by some feature such as a particular frequency. In this case, the signal has a single feature, a particular frequency, which varies along a single dimension, its intensity.

Even in this simple case, a receiver in natural situations faces a formidable problem. By the time the signal reaches the receiver, its intensity varies irregularly, as a result of variable attenuation and degradation of the signal during propagation. With some ingenuity and proper instruments, we can measure the intensity of the arriving signal in appropriate time intervals (for instance, the temporal resolution of the receiver's hearing). From this information, we can determine the probabilities of different intensities of the characteristic frequency as the signal reaches the receiver. These probabilities constitute the probability density function (PDF) for the intensity of that frequency during a signal.

At the same time, the receiver usually experiences background stimulation that can also include this characteristic frequency. For instance, this frequency might occur in other species' or individuals' vocalizations or in other environmental sources of sound, all irrelevant to the listener. Again, with some care we can determine the PDF for the intensity of this frequency in the background stimulation reaching the receiver.

If the distributions of intensities during the signal and background stimulation overlap, then the receiver (a listener) cannot avoid mistakes. Errors are inevitable whenever a receiver cannot completely separate signal and background. Only an observer with independent access to the source of the signal and the background can measure their properties separately. A receiver has no independent access to the signal. It must instead decide whether or not a particular intensity of the characteristic frequency merits response or not. Past experience with different intensities might lead to different expectations for the frequency of the signal and thus different levels of confidence in its decision to respond or not, but an isolated perception itself provides no basis for certainty.

A simple graph can introduce the issues that arise in this situation. We can plot the overlapping PDFs for intensity during the signal and background stimulation along the same axis (**Figure 1**). The subject's criterion for a decision is then represented by a threshold for response. In **Figure 1**, the PDFs are represented by normal distributions, with equal variances but different means. This simplified situation applies when background stimulation has a normal distribution of intensities and the signal has a fixed intensity, which is added to the background. Complications are addressed later, but they do not change the basic issues.

Once a threshold for response is chosen, then the total probability of a correct response (responding when a signal has occurred) is the integral of the PDF for signals from the threshold to infinity. The probability of a missed detection (failing to respond when a signal has occurred) is the integral of the same PDF from the threshold to negative infinity. Thus when a signal occurs, the probability of correct detection by the receiver equals one minus the probability of missed detection, $P_{CD} = 1 - P_{MD}$. Similarly, we can find the probabilities of false alarm (responding to background stimulation), PFA, and correct rejection (no response to background stimulation), P_{CR}, from integrals of the PDF for background stimulation. When only background stimulation occurs, $P_{FA} = 1 - P_{CR}$.

The first essential feature of signal detection is now apparent. As a receiver changes its threshold for response, the PFA varies with the probability of a correct detection (P_{CD}). By shifting the threshold for response to the right (toward higher intensities of the characteristic frequency), for example, a receiver can reduce its false alarms (responding when there is no signal present). Simultaneously, however, it increases its missed detections (not responding when a signal occurs). Clearly a receiver in this situation cannot simultaneously both minimize P_{FA} and maximize P_{CD}. This trade-off between correct detections and false alarms has fundamental implications for the evolution of communication (Wiley, 1994).

Another essential feature of signal detection is a distinction between the receiver's criterion for a response and the detectability of the signal. In this simple case, the receiver's criterion is represented by a threshold for response; the detectability of the signal is represented by the separation of the PDFs for signal alone and signal plus background (the difference between the means in relation to the standard deviation). A receiver's performance is determined by both of these variables.

When we present signals to animals, such as recordings of calls or songs, we often want to determine the subjects' attitude or responsiveness toward the signal. It is thus the subjects' criterion that interests us. In other cases, such as determination of sensory thresholds, it is the detectability of the signals that interests us. Signal detection theory allows us to separate the criterion for response from the detectability of signals. To see how, we can turn to a well-established application of this theory.

Signal Detection Theory in Experimental Psychophysics

The earliest application of signal detection theory to a behavioral problem was the determination of human sensory thresholds. Signal detection theory solved the problem of measuring the detectability of a signal despite differences in subjects' thresholds for responses. Procedures for this purpose are now well established (Green and Swets, 1966; Macmillan and Creelman, 1991; McNicol, 1972).

Before the application of signal detection theory, psychophysicists determined the absolute threshold for hearing sounds of a particular frequency by asking subjects to respond to faint sounds, barely separable from the background. The activity of auditory neurons in response to these sounds would barely differ from their spontaneous activity. These experiments confronted an insurmountable problem, because there was no satisfactory way to standardize the criteria different subjects used for responding, in other words, their thresholds.

Signal detection theory provides a solution to this problem by a simple modification of the experimental procedure. Subjects listen for a tone during brief intervals indicated by a cuing stimulus, for instance illumination of a light. During half of these intervals, selected at random, there occurs a tone of a particular frequency and intensity; during the remaining intervals there is no tone. The intervals with a tone allow an estimate of P_{CD}; those with no tone allow an estimate of P_{FA}. If the tone is loud enough, subjects detect the tone with high efficiency (high P_{CD} and low P_{FA}). If the tone is faint, this efficiency drops.

The subject's performance in this situation depends on both the detectability of the stimulus and the subject's criterion for response (in this case, a threshold). The literature in psychophysics often refers to a subject's criterion as a bias. For any constant level of detectability (the distance between the means of the two PDFs relative to the standard deviation), as a subject's threshold increases, P_{CD} increases as a function of P_{FA}. This function, called the *receiver operating characteristic* (ROC), increases monotonically from (0,0) to (1,1) in the unit square (**Figure 2**). As an exercise, try generating **Figure 2** from **Figure 1**, by varying the threshold for response.

To obtain an ROC, we must measure P_{CD} and P_{FA} at different thresholds for response. Psychophysicists use two basic methods. One involves direct manipulation of the subjects' thresholds, by rewards or instructions that place different weights on correct detections and false alarms. Another method involves asking subjects to rate their certainty for each response (for instance, 0 = absolutely certain no signal occurred, 10 = absolutely certain a signal occurred) (Egan et al., 1959; Macmillan and Creelman, 1991; McNicol, 1972). In the latter case, the experimenter uses different levels of certainty for different thresholds of response. For instance, for a high threshold, take all responses with certainty greater than 9 as positive responses for determining both P_{CD} and P_{FA} For a lower threshold, take all responses with certainty greater than 8, and so forth. Accuracy in estimating P_{CD} and P_{FA} at each threshold requires repeated tests for each subject.

The ROC then allows us to determine the detectability of a signal in a way that is independent of subjects' thresholds for response. As the detectability of a signal increases (the PDFs for signal alone and signal plus background move apart), the ROC moves away from the positive diagonal toward the upper left corner of the unit square, the point where performance is ideal ($P_{\mathrm{CD}} = 1$, $P_{\mathrm{FA}} = 0$). The closer the ROC approaches the upper left corner, the greater the detectability of the signal.

As the subject's threshold changes, on the other hand, its performance moves one way or the other along the ROC. As its threshold increases, a subject's performance approaches the origin ($P_{\mathrm{FA}} = 0$, $P_{\mathrm{CD}} = 0$). As its threshold decreases, its performance approaches the upper right corner ($P_{\mathrm{CD}} = 1$, $P_{\mathrm{FA}} = 1$). Thus changes in detectability of a signal shift the ROC away from or toward the diagonal, while changes in the subject's threshold shift its performance upward or downward along the ROC. Some study of **Figures 1** and **2** can clarify these relationships between the detectability of a signal, the threshold for response, and a subject's performance (its P_{CD} and P_{FA}).

Measurement of detectability is straightforward when the PDFs for background alone and for signal plus background are normally distributed with equal variance. The ROC in this case is symmetrical about the negative diagonal. If we plot the normal deviates or z-transforms of P_{CD} and $\mathrm{P}_{\mathrm{FA},}$ then the ROC is a straight line with unit slope (**Figure 3**), and the difference in z-scores,

$z(P_{FA}) - z(P_{CD})$, is the same for all points on this line. This difference, usually represented by d', represents the detectability of the signal. It equals the separation of the PDFs for background alone and signal plus background divided by their standard deviation. For alternative measures of detectability, all highly correlated with d', see discussions by Green and Swets (1966), McNicol (1972), or Macmillan and Creelman (1991).

Detectability is a measure of a receiver's ability to separate a signal from background stimulation; the analogous measure of ability to separate two signals is discriminability. The methods just described for measurement of the detectability of a signal also permit measurement of the discriminability of two signals. Instead of comparing responses to a signal and background stimulation, we compare responses to two signals in the presence of constant background stimulation.

General Assumptions of Signal Detection Theory

The theory of signal detection derives from assumptions about the nature of signals and their processing by receivers. This section considers these assumptions in order to establish the wide application of this theory. General assumptions are separated from some specific ones so that we do not discard the general theory entirely on the basis of questions about specifics. This section addresses general assumptions; the next considers specifics.

The theory accommodates a broad definition of a signal. Elsewhere, I have proposed that a signal is any pattern of energy produced by one individual (the signaler) and evoking a response from another individual (the receiver) without providing all of the power necessary to effect the response (Wiley, 1994). Some power is necessary to produce an alteration in the receiver's sensors, but the receiver itself provides essential power for the response. It is the necessary role of the receiver in producing a response that creates the essential elements of signal detection and, ultimately, all communication. Although a signal is similar to any stimulus that evokes a response, the term "signal" serves to emphasize the crucial importance of the limited contribution of power for the response.

The restriction of the sources and receivers of signals to living individuals (or their components) serves to include just those cases in which signalers and receivers might coevolve. This restriction is not essential; however, as signal detection theory addresses the optimization of a receiver's performance regardless of the source of the signals. Nevertheless, when both source and receiver are living organisms or their components, the possibility of coevolution raises particularly interesting issues, a topic we discuss later.

Signal detection theory also accommodates a broad scope for receivers. The two essential components of a receiver are a sensor and a mechanism for decisions. Each sensor is a perceptual channel tuned to a particular feature or dimension of stimulation (such as a particular band of frequencies of sound, a particular direction of a visual object, or a particular spectral-temporal pattern of sound). A decision to respond then depends on the output from one or more of these perceptual channels (**Figure 4**). Any channel is specified by its characteristic feature (for instance, the frequency of sound for maximal response from an auditory neuron) and its selectivity (often presented as its tuning curve or pass band).

Each channel produces an output that depends on the energy in its pass band within the broader range of energy impinging on the organism. This stimulation can include background energy of no interest to the organism (including irrelevant signals produced by other species or individuals and energy from the physical environment). The physiological mechanisms of channels often also produce spontaneous output. Consequently, a decision to respond based on the output of a channel often includes the possibility of false alarms and missed detections.

This model has broad generality (Green and Swets, 1966, Chapter 1). It is perhaps the most general model for an organism's responses to stimulation: a decision to respond or not depends on the output of a channel that receives combined signal and background. Green and Swets (1966) showed that the best rule for a decision to respond is a likelihood ratio that takes into account the expected frequencies of occurrence of signals. These basic ideas have a long history in psychology (Broadbent, 1958) and are familiar to ethologists and neuroethologists studying releasing mechanisms, stimulus filtering, and feature detectors.

The literature of psychophysics often contrasts "signal detection theory" with "threshold theory" (Green and Swets, 1966; Luce, 1963; Luce and Green, 1974; Macmillan and Creelman, 1991; Yonelinas, 2002). The distinction, however, is not fundamental. Threshold theory assumes some threshold above which a signal is always detected without error. Below this threshold, signals are detected with some fixed P_{FA} and P_{CD} (or some fixed ratio of these values). This theory thus requires at least two channels for the analysis of any one feature of a signal, one error-free for signals above the threshold, the other error-prone for signals below the threshold. These two channels, however, are equivalent to a single channel without normally distributed PDFs for background and for signal plus background (in this case the PDFs are rectangular; for full discussion, see Green and Swets, 1966; McNicol, 1972; Macmillan and Creelman, 1991). Only if we restrict the term "signal detection theory" to normally distributed PDFs with equal variance, are we forced to draw a distinction between this theory and "threshold theory." If we relax these restrictions, threshold theory becomes a special case of a general signal detection theory, based on a model of signal detection without restrictions on the distributions of outputs from perceptual channels.

Debate about these alternatives complicates much of the psychological literature on signal detection. In many cases, signal detection theory with additional assumptions of normality and equal variance can explain the properties of experimentally determined ROCs. The assumptions of normality and equal variance are best approached by examining the procedures for measuring detectability.

Specific Assumptions of Signal Detection Theory: Measuring Detectability

Calculation of *d*' from a single pair of measurements and determinations of absolute sensory thresholds requires some specific conditions: (1) normally distributed PDFs with equal variance; (2) fixed criteria for responses; and (3) cuing of responses. This section considers each of these requirements. Although each is critical in special cases, none is necessary for measurements of detectability in general.

1. Normal Distributions with Equal Variance

For sensory discriminations under laboratory conditions, the relevant PDFs are often nearly normal with nearly equal variances. The clearest evidence is an ROC symmetrical around the negative diagonal in the unit square and linear in probability space (with z-transformed axes for P_{CD} and P_{FA}) with slope equal to 1 (Green and Swets, 1966; McNicol, 1972). In this case, $d' = z(P_{\mathrm{FA}}) - z(P_{\mathrm{CD}})$ provides an unambiguous measure of detectability.

If the PDFs are not normally distributed or have unequal variances, then the picture changes. If variances are not equal, the ROC lacks symmetry around the negative diagonal. When plotted in z-transform space, the ROC has a slope equal to the ratio of variances. If the PDFs are not normally distributed, the ROC changes shape and is no longer linear in z-transform space. Consequently, when either normality or equal variance is violated, $d' = z\ (P_{\mathrm{FA}}) - z(P_{\mathrm{CD}})$ makes little sense as a measure of detectability.

When normality or equal variance does not apply, we must use an alternative measure of detectability. A simple one is the area between the ROC and the positive diagonal of the unit square. This area measures the displacement of the ROC away from the positive diagonal and toward the point of maximal performance at the upper left-hand corner; *d*′ provides a measure of this displacement only for a symmetrical ROC.

2. Optimal Criteria

Accurate measurement of absolute sensory thresholds requires that subjects use an optimal criterion or rating scale for any set of experimental conditions. Variation among subjects, or variation among trials for any one subject, results in an underestimate of d' for maximal performance and also an underestimate of any difference in variances between signal and background. In carefully conducted psychophysical experiments, these possible errors turn out to be slight (Macmillan and Kaplan, 1985; McNicol, 1972, pp. 202–204).

This assumption that subjects use an optimal criterion is less critical for an investigation of communication, when an organism's actual performance has greater interest than its maximally possible performance. In this case, we can combine observations from different subjects by averaging z-scores to obtain a composite value of d' (Macmillan and Kaplan, 1985; McNicol, 1972, p. 112). If subjects' criteria or ratings vary, these composite measurements of detectability reflect expected average performance. Alternatively, we could study each individual's ability to detect or to discriminate signals.

3. Cuing of Responses

Any measurement of the detectability of a stimulus requires null (background only) presentations, which permit measurement of P_{FA}, the probability of response without the signal present. In laboratory experiments, a cuing stimulus identifies intervals in which the subject must make a decision. This procedure assures equal decisions with and without the signal present. In field experiments this device is not possible. However, we can still include null presentations with no stimulus; even better, white noise or prerecorded background sounds might serve as a null stimulus. Alternatively, one could abandon attempts to measure the absolute detectability of any one stimulus and consider only the discriminability of two signals. In this case, a balanced experimental design could include equal numbers of presentations of the two signals.

The absence of null presentations confounds interpretation of a large body of research on human vigilance (Davies and Parasuraman, 1982; Mackie, 1977). Studies of vigilance and field studies of responses to playback have some similarities. In both cases, subjects experience long intervals between infrequent occurrences of a stimulus. The long periods without signals inevitably make $\mathrm{P_{FA}}$ very small during any brief interval when the signal is absent. Consistently small $\mathrm{P_{FA}}$ makes a meaningful ROC difficult to construct. Despite some suggestions for ways to circumvent this problem (Egan et al., 1961a; Watson and Nichols, 1976), there seems to be no convincingly satisfactory solution. When we cannot measure false alarms, by means of cuing, null presentations, or comparisons of two signals, determination of an ROC is problematic. Measurement of $\mathrm{P_{FA}}$ is essential for a full understanding of a receiver's performance. A later section discusses some practical possibilities for solving this problem in field studies of animal communication by means of playbacks.

The two general results of signal detection theory—the interdependence of P_{CD} and P_{FA} and the distinction between the receiver's criterion and the detectability of the signal—do not depend on the specific assumptions of normality and equal variance and are not affected by the practical difficulties of measuring detectability or discriminability. These two general features of signal detection are alone sufficient to clarify the determinants of a receiver's performance.

Properties of Signals that Affect a Receiver's Performance

Signal detection theory makes it clear that any receiver's performance in detecting or discriminating signals has limits. Furthermore, these limits are in part determined by properties of the signals. Predictions about these determinants of a receiver's performance have in many cases been repeatedly confirmed by psychophysical studies of humans, but the results of these studies have broad application to signal detection in general and thus to all forms of communication. Consider three properties of a signal that influence a receiver's performance: (1) contrast, (2) redundancy, and (3) uncertainty. We shall see that the inevitable effects of these three properties of signals explain a lot of "receiver psychology."

1. Contrast

Contrast and detectability are so closely related that it requires care to distinguish them carefully. As explained earlier, detectability is the difference between the means, in relation to the standard deviations, of background alone and background plus signal in the output of some perceptual channel (for instance, in the responses of an experimental subject). Contrast is an analogous difference in the stimulation at the input to a channel (in the stimulation impinging on the subject). Unlike detectability, contrast depends only on the properties of the external stimulation reaching an organism and not on the properties of the organism's perceptual channels.

Contrast usually increases detectability. The influence of contrast on a subject's performance is so clear that it has received little explicit study by psychophysicists. One such study, included in one of the first applications of signal detection theory to perception, showed that log d' increased linearly with log intensity for a signal in the presence of constant background stimulation (Tanner and Swets, 1954).

Because we define contrast by the properties of a signal in relation to the background stimulation impinging on an organism, detectability of the signal depends on both its contrast and the selectivity of the perceptual channel. This dual determination of detectability is the basis for a procedure in psychophysics for determining bandwidths of sensory channels. In the case of hearing, the intensity of broad-spectrum background sound (white noise) that can mask a signal of a particular frequency depends on the bandwidth of the auditory channel. In fact, the signal-to-noise ratio (a measure of contrast) for complete masking of a single frequency with broad-spectrum noise equals the effective bandwidth of the auditory channel for that frequency.

The dual determination of detectability implies that the intensities of signals and background stimulation impinging on an organism do not alone allow us to predict an organism's performance. For instance, the intensity of a particular frequency of sound, or hue of light in a signal, and in the background are not enough to allow us to predict the detectability of that sound or light for a particular organism. To determine the influence of contrast on detectability, we must study the organism's responses, at either the neural or behavioral levels.

Study of contrast and detectability in natural situations is still rudimentary (Klump, 1996). For instance, despite many studies of sound propagation in natural environments and its influence on the evolution of bird songs (reviewed by Naguib and Wiley, 2001; Wiley, 1991), we know little about the properties of background sound in relation to acoustic signals in natural situations. Such studies of acoustic contrast would require recordings of signalers with omnidirectional microphones at typical positions for conspecific listeners. To extend these studies to detectability would require adjustments for the directionality and selectivity of the listeners' hearing.

Only one study has shown how background noise affects the detectability of acoustic signals in natural situations. Measurements of auditory thresholds in great tits *Parus major*, in the absence of noise, reveal greatest sensitivity to frequencies between 2 and 4 kHz, lower than most of this species' vocalizations. However, critical bandwidths remain nearly constant over a wide range of frequencies up to 8 kHz. Consequently, in the presence of wind in a forest, which produces noise decreasing exponentially in intensity with increasing frequency, the frequency for greatest detectability shifts to 8 kHz (Langemann et al., 1998). It is also clear the birds and mammals can increase the intensity of vocal signals in the presence of background sound, presumably to improve the contrast of their signals with the background (Brumm, 2004;

Brumm and Todt, 2002; Brumm et al., 2004; Cynx et al., 1998; Leonard and Horn, 2005). Shifts in frequency to increase contrast with background noise are not so well documented. The clearest case is again the great tit, which uses higher dominant frequencies in its songs in urban environments with predominantly low-frequency noise (Slabbekoorn and Peet, 2003).

Contrast and detectability of visual signals is more complex. Unlike acoustic signals, for which the signaler generates the power to produce the signal, visual signals usually rely on reflectance or scattering of light from other sources. As Endler (1990, 1993) explains, the spectrum of light arriving at a receiver's eyes from an object depends on the product of the irradiance spectrum, the reflectance spectrum of the object, and the transmission spectrum (the spectra of the incident, reflected, and transmitted light, Q, R, and T). The contrast between a visual signal and its background thus depends on the contrast between QRT for the signal and the background. Q, which depends on the photic properties of the environment, can vary substantially with microhabitat (Endler, 1993; Gomez and The´ry, 2004). These principles apply to male manakins, small birds that use bright colors in their plumage to produce visual displays at leks in the understory of tropical forests. Both the reflectance spectra of patches in their plumage and the placement of their leks in the forest serve to increase the contrast of their displays with the visual background (Endler and The´ry, 1996; Heindl and Winkler, 2003). Furthermore, Uy and Endler (2004) have shown that, in one species, males increase the contrast of their plumage with the background by clearing fallen leaves from their display sites. Contrast between different parts of a signal is also affected by choice of location (Endler, 1993; Heindl and Winkler, 2003), but this within-signal contrast is a form of structural redundancy, discussed in the next section.

One consequence of the dependence of visual signals on environmental irradiance is that changes in habitats can drastically alter contrast of signals with background. A case in point are the numerous endemic species of cichlids in Lake

Victoria. Many of these recently evolved species differ mainly in male coloration and mate choice by females. Increased turbidity of some parts of the lake in recent decades, as a result of sedimentation from human activities, is associated with a loss of many species (Seehausen et al., 1997). Apparently, the species-specific colorations of the males no longer contrast enough to allow females to differentiate them.

Contrast applies to complex signals as well as to signals with a single characteristic feature. As with simpler signals, there has been little investigation of complex signals in the presence of background stimulation. One exception is human speech. Early experiments showed that human subjects have trouble understanding one person speaking in the presence of others, the so called "cocktail-party problem" (Cherry, 1953; Cherry and Taylor, 1954). Similar tasks requiring discrimination of one conspecific's vocalizations from those of other conspecifics in the background recur in many natural situations, for instance in choruses of frogs or insects, colonies of seabirds, and dawn choruses of birds or primates. Detection and discrimination in these situations have received little attention. One such study in a colony of king penguins (*Aptenodytes patagonicus*) confirmed that the presence of large numbers of conspecifics increased attenuation and degradation of the adults' calls that allow chicks to recognize their parents (Aubin and Jouventin, 1998). The situation is particularly difficult because the noise has nearly the same spectral distribution as the signals of interest to a chick. Nevertheless, these chicks can recognize their parents' calls even when the overall signal-to-noise ratio is less than 1. In such "cocktail-party" situations, birds as well as humans use cues for spatial localization to increase the effective signal-to-noise ratio of signals in more evenly distributed noise (Cherry, 1953; Cherry and Taylor, 1954; Dooling, 1982). In this case, contrast between signals consists mostly of differences in location.

2. Redundancy

Redundancy results from predictable relationships between the parts of a stimulus, either in time or space. It takes two forms, both of which improve detectability of a signal. Sequential redundancy consists of fixed temporal relationships between components of a signal. Repetition of a signal, the simplest form of sequential redundancy, increases its detectability (Swets and Birdsall, 1978; Swets et al., 1959). In fact, the detectability of tones increases with the square root of the number of presentations. This result is consistent with an assumption that each instance of a stimulus is assessed independently (Swets et al., 1959).

All psychophysical experiments on detectability use an alerting signal to tell the subject when to respond. An alerting signal, one with high contrast and low uncertainty, accompanying a more informative signal is a special case of redundancy. Although many natural signals might include alerting components (Richards, 1981a), this possibility has received little attention. Simultaneous redundancy consists of fixed relationships between concurrent dimensions of a signal. Simultaneous redundancy can take the form of multiple components with fixed spatial relationships in a visual signal, multiple molecular components in an olfactory signal, or multiple components with fixed spectral relationships in an acoustic signal. Such a multidimensional stimulus is more detectable than one with a single feature. The increase in detectability with the number of features characterizing a stimulus again suggests that human observers assess each feature independently (Macmillan and Creelman, 1991; Mulligan and Shaw, 1980; Shaw, 1982). An interesting twist on redundancy involves predictable relationships within the background noise rather than within the signal of interest. If different frequencies in noise are subject to synchronized amplitude modulation (called comodulation), then it is possible to use the properties of noise in one band of frequencies to improve detection of a signal in another band. This "comodulation masking release" has been demonstrated in both humans and birds (Klump and Langemann, 1995; Langemann and Klump, 2001; Nieder and Klump, 2001).

3. Uncertainty and Unfamiliarity

Uncertainty about signals takes two forms, each of which decreases detectability. Intrinsic uncertainty occurs when a subject lacks prior information about a signal's features, including the interval of time and location in which it might occur. Extrinsic uncertainty occurs when a subject must respond to many different signals. A subject can have prior information about the features of each signal but still face uncertainty about which signal will occur. Multiplicity of signals reduces the detectability of each of them.

Uncertainty about the features of signals reduces their detectability (Green, 1961; Pelli, 1985). Detectability also decreases when observers are uncertain about the locations or intervals of time in which signals might occur (Egan et al., 1961b; Watson and Nichols, 1976; Starr et al., 1975; Swensson and Judy, 1981). These latter situations are in fact special cases of the detection of signals with uncertain features.

Uncertainty about which of several signals might occur also reduces their detectability. For instance, if human observers are asked to report any of several possible signals, the overall detectability of the signals decreases as the number of alternatives increases (Cary and Reder, 2003; Nolte and Jaarsma, 1967).

Human performance in detecting multiple signals again implicates independent perceptual channels. It is as if a separate channel assesses each signal's characteristic feature, and the subject decides that a signal has occurred when the criterion in any channel is met (Cohn, 1978; Green and Birdsall, 1978). These conclusions rest on a comparison of the ROCs of subjects detecting different numbers of signals. This analysis also confirms that the reduction in detectability of signals in this situation results from the uncertainty of the task, not from any change in the observers' criterion for response. Thus detection of signals from a repertoire of possibilities is inherently more difficult than detection of a single signal specified in advance.

Unfamiliarity also makes signals more difficult to detect. For instance, the frequencies of words in common usage influence their thresholds for visual recognition (Pierce, 1963). Other studies have confirmed that high- frequency words are more detectable than low-frequency ones (although memory of high-frequency words presented previously is less accurate) (Broadbent, 1967; Glanzer and Adams, 1985; Glanzer et al., 1993; Pollack et al., 1959). Thus greater familiarity with a stimulus increases its detectability, just as greater uncertainty reduces it.

Human performance during vigilance fits the same pattern. The greater the uncertainty about the features, timing, or location of possible signals, the lower the efficiency of the observer (Davies and Parasuraman, 1982; Davies and Tune, 1970; Loeb and Alluisi, 1977; Warm, 1977). Studies of vigilance have not provided definitive evidence that detectability changes, as opposed to the subject's criterion, because such studies, as explained earlier, do not allow analysis of the ROC. Nevertheless, these results resemble those of studies with a full analysis of detectability and thus reinforce the conclusion that uncertainty about a stimulus, in any form, reduces its detectability.

Classification of Signals in Addition to Detection

Although in many situations it is reasonable to assume that an animal's task involves no more than detection of an appropriate signal, in others some classification of a stimulus is essential. Detection, for example, is involved when an individual responds to a suitable mate or to its offspring or chooses a diet based on profitability of prey. Classification, on the other hand, is required when it recognizes several social partners or chooses a diet with an optimal mixture of nutrients. The discussion so far has focused on detection of a signal in noise. This section considers the use of signal detection theory to understand a receiver's performance when classification is as important as detection.

An experiment to show detection of a signal is designed so that the subject must make a binary decision about the occurrence of the signal, “yes” or “no,” go or no-go. An experiment to show discrimination likewise requires only a single binary decision, either “signal 1” or “signal 2.” Other situations, however, require both detection and subsequent classification of signals. Detection plus classification requires one of at least three responses (“no,” “1,” or “2”) as a result of at least two binary decisions (“yes” or “no;” if “yes” then “1” or “2”).

Detection plus classification is the basis for recognition or identification, as these terms are often used. In some discussions, however, recognition means detection of multidimensional signals or detection of signals with uncertain features, situations discussed in the previous section. These cases require single binary responses to a multiplicity of possible signals. The distinguishing feature of a classification of signals, in contrast, is the multiplicity of possible responses.

A few experiments confirm that classification in addition to detection is a more difficult task for receivers than detection alone. For instance, the task of identifying a stimulus as familiar or not requires less attention during previous exposures to the stimulus than does recollecting specific associations of a stimulus (Dobbins et al., 2004). Female frogs (*Hyla ebraccata*) detect a conspecific male’s calls in background noise from a natural chorus at signal-to-noise ratios above 3 dB. Yet they express a preference for those calls with lower fundamental frequencies only at signal/ratios greater than 9 dB (Wollerman and Wiley, 2002). At intermediate signal-to-noise ratios, females did not discriminate between otherwise preferred and nonpreferred males’ calls, even though she could detect these calls.

Classification in addition to detection has surprisingly complex influences on a receiver’s performance. To analyze these complexities and to assess their influence on receivers, we first consider a basic experiment. This approach leads to more complex ones and ultimately to a theoretical justification for a general principle: a receiver’s performance in a task requiring classification is inevitably lower than in a comparable task requiring only detection.

To investigate detection plus classification, an experiment might present background alone and background in combination with each of two signals. With human subjects, we can simply ask for two responses, first "yes" or "no" for the presence of a stimulus, then "1" or "2" for the class of stimulus, provided one has been detected. Because classification presupposes correct detection of signals, the probability of correct classification can never exceed the P_{CD}. Some evidence for "subliminal" classification does not alter the situation significantly (Macmillan and Creelman, 1991, p. 255).

One approach in a study of this sort is to calculate both an ROC and an analogous *identification operating characteristic* (IOC). To construct the ROC for this case, one measures P_{CD} as the probability of a correct "yes" response when either signal occurs and P_{FA} as the probability of a "yes" response when no signal occurs. For the analogous IOC, one measures P_{CD} as the probability of correct identification of a signal when it occurs; P_{FA} is still the probability of a "yes" response when no signal occurs (Benzschawel and Cohn, 1985; Green and Birdsall, 1978; Green et al., 1977; Macmillan and Creelman, 1991). The IOC, thus defined, can be derived from the ROC for detection of uncertain signals discussed earlier. Despite this theoretical advantage, the IOC fails to consider errors of classification once a signal is detected and thus provides an unrealistic measure of a receiver's performance.

A better approach in a study of detection plus classification is to consider a bivariate plot of PDFs, with one axis for a measure of the characteristic feature of each stimulus (**Figure 5**). If the characteristic features of the two signals are orthogonal (in other words, if they vary independently), the PDFs for background only and for each signal in combination with background lie along two perpendicular axes. A receiver's performance then depends on three thresholds: two that separate background from each signal in combination with background (T1 and T2) and a third that separates the two signals (T3, **Figure 5**). This third threshold differentiates the two signals based on the ratio of measures of their respective characteristic features. The slope of threshold T3 changes, as the receiver alters its criterion for classifying the signals.

This experiment thus allows measurement of three d' values (Macmillan, 2002; Macmillan and Creelman, 1991; Tanner, 1956): between background (B) and background plus one of the signals (B + S1), between B and B + S2, and between B + S1 and B + S2. Suppose the receiver processes the characteristic features of the two signals independently, as predicted for orthogonal features, and the variances of the three PDFs are equal, as predicted for constant signals added to background, with equal variance in each signal's characteristic feature. These three d' values have a Pythagorean relationship, $d_1 = \sqrt{(d_2^2 - d_3^2)}$ as seen by geometry in **Figure 5**, in which each d' is proportional to the distance between the means of the respective PDFs.

An even more robust experiment would include a fourth stimulus, background in combination with both signals at once, B + S1 + S2. The six d' values in this case specify the nature of any interaction in processing the features of the two signals (masking of one signal by the other, inhibitory interaction between channels, correlation of the background in the two channels) (Klein, 1985; Thomas, 1985).

To understand the consequences of detection plus classification for a receiver's overall performance, we can compare P_{CD} and P_{FA} for detection plus classification with those for simple detection. The probability of correct response to a particular signal (P_{CD} for detection plus classification) is always less than or equal to that for simple detection. As the threshold for classification, T3, decreases in slope, P_{CD} for detection plus classification increases from near 0 to a value approaching P_{CD} for simple detection (**Figure 5**).

The situation for P_{FA} is more complex, because it involves two kinds of false alarm responding when only background occurs or when the alternative signal occurs. Because classification must follow detection, the two kinds of false alarm are not independent. Consequently, to combine the P_{FA} for simple detection and the P_{FA} for detection plus classification requires information about the relative frequencies of these two situations. A full analysis of this situation is not yet available.

Analysis of this situation is simplified by considering only false alarms for detection. False alarms in this narrow sense include only responses to background stimulation and thus include only false alarms for detection and exclude those for classification. For any level of false alarm in this narrow sense, classification in addition to detection reduces correct responses to signals in comparison to detection alone. Conversely, for any level of correct detections of signals, classification in addition to detection increases false alarms in this narrow sense (Macmillan, 2002; Starr et al., 1975). Classification plus detection, in comparison to simple detection, thus inevitably results in more false alarms by a receiver, even in the narrow sense. Classification thus inevitably reduces a receiver's performance in comparison to detection alone.

Complex Patterns: Extension of the Concept of Channels

Signal detection theory, as we have seen, describes decisions based on the outputs of perceptual channels. Detection and discrimination, the focus of discussion so far, suggest that the perceptual channels under consideration are sensory receptors and their immediate neural connections. Peripheral mechanisms of perception have been the main concern of many applications of signal detection theory, especially in studies of hearing.

Nevertheless, the theory applies equally well to more cognitive aspects of nervous systems. A channel can in fact represent any step in the hierarchy of perceptual analysis of a signal. It could represent "detection" of a species-specific song, for instance, when the issue is not whether or not a listening bird can hear each of the component frequencies but whether or not the entire pattern fits some criterion for a decision to respond.

Such pattern detection has all the same general properties as feature detection. A channel for pattern detection produces an output that reflects the presence of components with particular sequential or simultaneous relationships. Irrelevant background stimulation can include similar relationships, differing in unpredictable ways from those in the signal, and the mechanism of the channel can itself include some unpredictability. A criterion for a decision to respond based on the output from such a pattern-detecting channel inevitably results in false alarms and missed detections, just as from a feature-detecting channel.

Thus all of the preceding discussion of signal detection theory applies equally well to complex, as well as simple, perception. It applies to recognition of conspecific songs, to recognition of the vocalizations of mates, offspring, or neighboring individuals, to mate choice based on complex repertoires, and to interpretation of subtle innuendos in the close-range vocalizations of group-living animals—signal detection theory applies to all communication.

Evolution of Signaling and Reception

Signal detection theory suggests ways that receivers and signalers could coevolve (Wiley, 1994). We can understand many features of this coevolution by applying principles of signal detection first to optimizing receivers' performance and then to optimizing signalers' behavior. Because receivers provide the essential power for responses, their adaptation is primary. Nevertheless, signal detection theory shows that receivers do not necessarily get what they want. Because of the inevitable limitations on their performance, receivers can attain optimal, but not ideal, performance. Signalers can then evolve in response to the conditions set by their intended receivers. If changes in signalers' behavior alter the features or frequency of signals, receivers might evolve new optima for their own performance. Then signalers might evolve new features of signals. It seems probable that this form of coevolution could either reach an equilibrium or propagate perpetual lags between the adaptations of signalers and receivers.

Receivers can optimize the net utility of their decisions to respond or not by adjusting their criteria for response. The net utility for a receiver's decision depends on the probabilities and payoffs (net gains, positive or negative) of correct detections, missed detections, false alarms, and correct rejections (for details, see Wiley, 1994). The payoffs from these four possible outcomes must be measured in units relevant to natural selection. The probabilities of these outcomes, we have seen, depend on the discriminability of signals and the receiver's criterion. Depending on these payoffs and probabilities, the optimal criterion for a receiver can lie anywhere between adaptive gullibility (a low criterion for response when missed detections are especially costly) and adaptive fastidiousness (a high criterion for response when false alarms are especially costly). Gullability of receivers should result in the evolution of dishonest signals, fastidiousness in the evolution of exaggerated signals (Wiley, 1994).

Signaling should evolve to increase the predictability of responses from intended receivers. As a result, signals should often evolve to improve detectability (Wiley, 1983, 1994), so receivers can in turn evolve criteria that permit high P_{CD} and low P_{MD}. Greater contrast and redundancy and less uncertainty about a signal's features, including its timing and location, all increase detectability and thus ultimately increase the probability of responses by receivers. Signal detection theory can explain why these properties of a stimulus affect detection and consequently learning and memory (the "receiver psychology" of Guilford and Dawkins, 1991, 1993). The widely reported phenomenon of peak shift in discrimination learning (Enquist and Arak, 1993; Guilford and Dawkins, 1991, 1993; ten Cate and Bateson, 1988; Weary et al., 1993) follows from maximizing the net utility of a receiver's criterion for response (Lynn et al., 2005). When false alarms are more costly than missed detections, it pays for a receiver to adapt a strict criterion for response. Because the adaptive solution is to respond to extremes of signal properties in one direction rather than the other, in order to minimize false alarms, peak shift is the result.

On the other hand, unintended receivers (eaves-dropping predators and parasites or conspecific rivals, for instance) can reduce the advantages of highly detectable signals. Properties that improve detectability, such as redundancy and predictability, also limit possibilities for encoding of complex information, which requires variation rather than constancy in signals (Wiley, 1994). Signals might thus evolve a compromise between advantages of detectability and advantages of privacy or complex coding.

Interpretation of Playback Experiments with Signal Detection Theory

Experimental studies of communication depend on presentations of signals to subjects in order to record their responses. Signal detection theory suggests new approaches for designing and interpreting such experiments. First of all, it calls into question the use of clear signals. Because the ability of animals to detect or to discriminate any signals depends on background stimulation, experiments with intense signals and weak background stimulation often have little relevance to communication in natural situations. Signal detection theory, however, does not simply suggest cautious interpretation of playback experiments. It also identifies two distinct reasons why results should depend on background stimulation: both the features of effective signals and a receiver's criterion for response should change with the level of background stimulation.

Many investigations of the features of signals that make them effective in eliciting responses have employed clear signals and minimal background stimulation. This approach is unlikely to provide a full understanding of communication because, as the preceding review has indicated, the features of effective signals, those that optimize receivers' performance, differ in the presence of high and low background stimulation. Signals effective when background stimulation is low could prove much less so when background stimulation is high. In the latter case, we should expect greater emphasis on features that contribute to detectability of signals (contrast, redundancy, low uncertainty, familiarity). Experiments with playbacks have so far never considered the possible effects of background stimulation on detectability of signals.

The interpretation of responses is also complicated by the possibility of confounding detectability of signals with criteria for responses. In studies of animal communication, experiments are usually interpreted in terms of the subjects' attitude toward the experimental signals. For instance, do subjects have a lower threshold for a particular response to one type of signal in comparison to another? Yet the probability of response depends both on the listener's attitude (its threshold or criterion) and on the level of the signal in relation to background stimulation as perceived by the listener (the detectability of the signal). A few studies of responses to bird songs in the field have considered both of these possibilities (Brenowitz, 1982; Richards, 1981b), but all have so far relied on indirect evidence to separate them. Even differences in responses to loud, repeated, clean signals might reflect differences in detectability of signals rather than differences in receivers' criteria for response. When it is important to be sure that the receivers' attitude (criterion) differs, only an ROC analysis can separate these possibilities.

Signal detection theory also shows how to characterize the general properties of perceptual channels by comparing responses to at least three types of signals. Each pair of signals elicits responses that depend on outputs from a perceptual channel or combination of channels. Although only neurophysiology can determine the neural components and mechanisms of these channels, we can nevertheless learn something about their overall properties even without knowing the details of their mechanisms. For instance, are the pattern-detecting channels for each of the three possible pairs of signals independent (A-B, B-C, A-C)? Measuring the discriminabilities for the three possible pairs of signals can provide an answer. As explained earlier, discriminabilities that summed would indicate completely shared channels; discriminabilities with Pythagorean relationships would indicate completely independent channels; intermediate results would suggest partially correlated channels.

Practicalities of Experiments in Natural Situations

To take advantage of these possibilities, we must measure detectabilities and discriminabilities in the field. To accomplish this task, we have to broaden the way we think about experiments with playbacks. Presentation of loud, repeated, clear signals close to subjects provides little information for comparisons of detectability or discriminability of signals. Instead, for this purpose, it would be better for each trial to present a brief (perhaps a single) stimulus in combination with background stimulation. Furthermore, the nature of the background must become part of the experimental design.

To determine the detectability of a single stimulus, we can use background stimulation as a null stimulus (background only) for comparison with the signal (background plus signal). To determine the discriminability of two signals, the problem of a null stimulus does not arise. Nevertheless, including a null stimulus in the experimental design adds the possibility of a full analysis of detection plus classification, as described earlier. An ROC can then allow evaluation of normality and variance in the outputs of the channels involved and thus choice of an appropriate measure of detectability or discriminability.

To construct an ROC from field studies of animals, a rating scale is likely to be the method of choice. To do so, we must first determine the distribution of some measure of response (perhaps the first principal component of all behavioral measures) across all trials. Depending on sample sizes, we can assign scores, for instance, to quartiles or deciles of this distribution. These scores provide nonverbal ratings of the subjects' levels of confidence in discriminating between the two signals. The distributions of scores for each signal then generate pairs of P_{CD} and P_{FA} for the construction of an ROC.

A practical problem in measuring ROCs in the field is the limited number of trials. Experiments with animals in the field can rarely expect, as psychophysical experiments do, to present signals hundreds of times to each subject and then to examine each subject's ROC separately. Field studies will probably have to combine data from different subjects and thus determine only characteristics of populations, ideally ones as homogeneous as possible. Nevertheless, practical numbers of trials could yield useful measures of detectability in experiments with rating scales (McNicol, 1972, Chapter 5).

Once an ROC is constructed, we can apply standard procedures for calculating detectability (or discriminability) of the signals. Furthermore, each pair of scores used to construct the ROC reveals the subjects' average criterion under particular conditions. Procedures for calculating detectability or discriminability from a rating scale, summarized earlier, are thoroughly reviewed by McNicol (1972). The location of the criterion for response under particular conditions is best specified by its absolute location with respect to the underlying PDFs. Macmillan and Creelman (1990) recommend simple measures, such as $(P_{\mathrm{CD}} + P_{\mathrm{FA}})/2$ or $-[z(P_{\mathrm{CD}}) + z(P_{\mathrm{FA}})]/2$. With these procedures, the application of signal detection theory to field studies of animal communication seems unlikely to encounter insurmountable problems.

Summary

Signal detection theory involves a level of abstraction unfamiliar in field studies of animal communication. Mastering its implications, however, leads to some strong predictions about the evolution of signals and responses and to some new procedures for investigating animal communication.

A consequence of this approach to communication is the fundamental conclusion that a receiver cannot independently adjust its P_{CD} and P_{FA}. The only exception is the limiting case in which the output of a channel in the presence of a signal is perfectly distinct from the output in its absence, so $P_{\mathrm{FA}} = 0$. Otherwise, no matter how the criterion for response changes, any change in P_{CD} is accompanied by a corresponding change in P_{FA}.

This compromise leads ultimately to a prediction that receivers evolve to optimize the net utility of their responses.

The optimum might lie anywhere between extremes of gullibility or fastidiousness. In turn, signalers should evolve to balance the often incompatible advantages of increased detectability of signals, increased complexity of encoding, and restriction of signals to intended receivers.

A second consequence of signal detection theory is the fundamental distinction between the detectability of a signal and the receiver's criterion for a response. Detectability depends on the contrast of the signal impinging on the subject and on the selectivity of the subject's perceptual channels. A receiver's criterion for response depends on its attitude toward the output of its perceptual channels, as a result of a decision to accept particular P_{FA} and P_{CD}.

Because any receiver's responses to stimulation depend on both detectability of the stimulus and criterion for response, a definitive interpretation of responses requires attention to both. For a full interpretation of a receiver's performance, it is necessary to include null presentations in experiments in order to measure false alarms as well as correct detections.

Signal detection theory thus suggests new ways to design and to interpret experiments that compare responses to stimulation. Although some practical difficulties face any application of signal detection theory to field studies, none seems insurmountable.

With this approach, we stand to learn more about (1) the adaptations for communication in situations with high background stimulation, such as in choruses or complex social groups or at long range, (2) the effects of contrast, redundancy, reduced uncertainty, and familiarity on receivers' abilities to detect and discriminate signals, and (3) the evolution of exaggeration or dishonesty in signals as a consequence of the evolution of receivers' performance. In all of these ways, signal detection theory can advance our understanding of both the physiology and the evolution of communication.

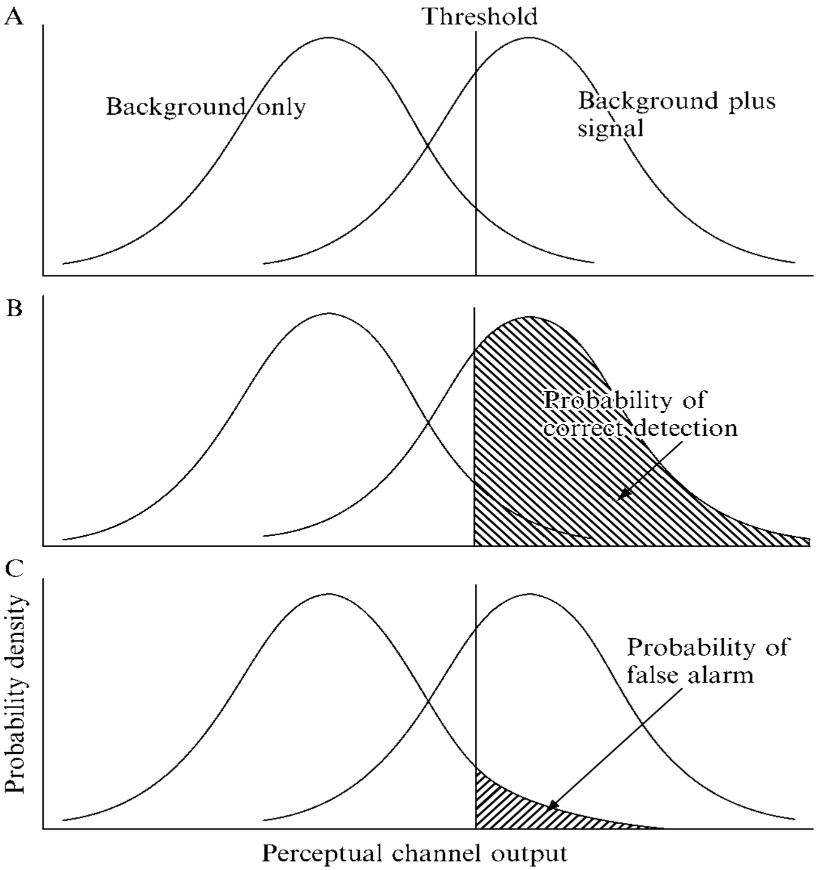


Figure 1. The basic situation described by signal detection theory. (A) The levels of back- ground stimulation with and without a signal are represented by the outputs from a perceptual channel. The probability of an output as a function of the level of the output is a probability density function, PDF, for the output. A decision to respond involves selecting a criterion (in this case, a threshold in the output of the channel above which a response occurs). (B) Any such threshold results in a probability of correct detections, P_{CD}, the area under the PDF for background plus signal to the right of the threshold. (C) Any threshold also results in a probability of false alarms, P_{FA}, the corresponding area under the PDF for background alone.

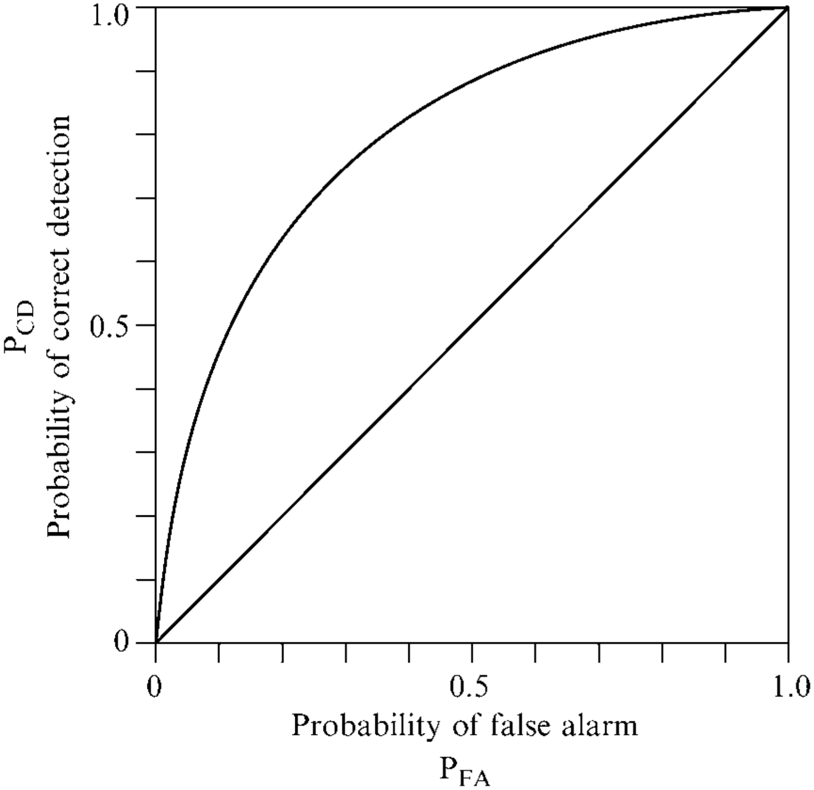


Figure 2. A receiver operating characteristic (ROC) results from plotting P_{CD} as a function of P_{FA}, as the threshold for response varies. The ROC is symmetrical about the negative diagonal of the unit square provided the two PDFs have normal distributions and equal variances. The separation of the means of the PDFs determines how far the ROC lies from the positive diagonal and thus how nearly it approaches the point of ideal performance, the upper left- hand corner. This illustration shows the ROC when the means are separated by one standard deviation ($d' = 1$).

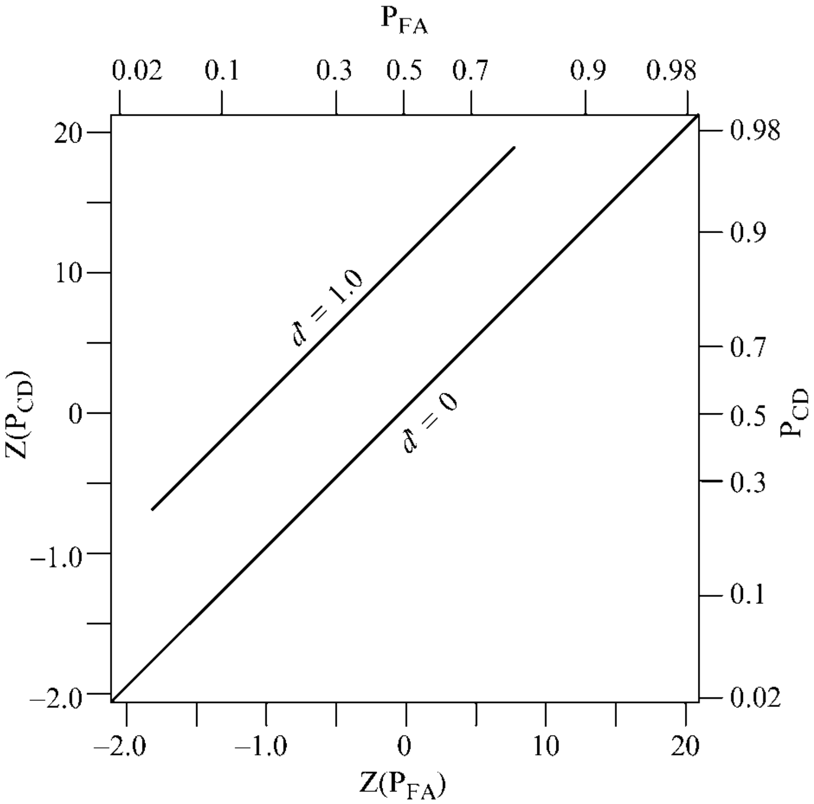


Figure 3. An ROC plotted on probability (z-transformed) axes is a straight line with slope = 1 in the case of normally distributed PDFs with equal variance. This illustration shows the same ROC as **Figure 2**.

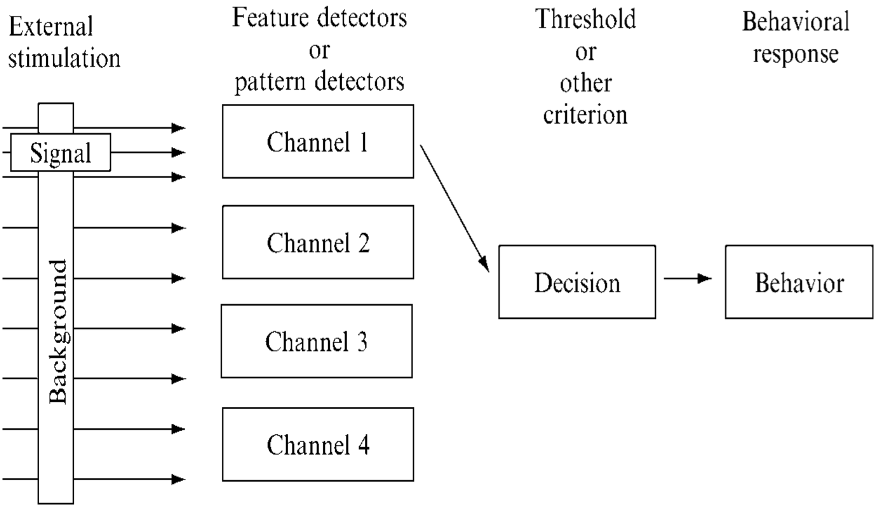


Figure 4. The general model for signal detection involves perceptual channels that analyze features or patterns in stimulation impinging on the receiver. The output of one or more channels forms the basis for a decision to respond (in the form of a multidimensional criterion for response). Channels and decisions might represent distinct neurons or populations of neurons, or a single neuron might combine these two properties.

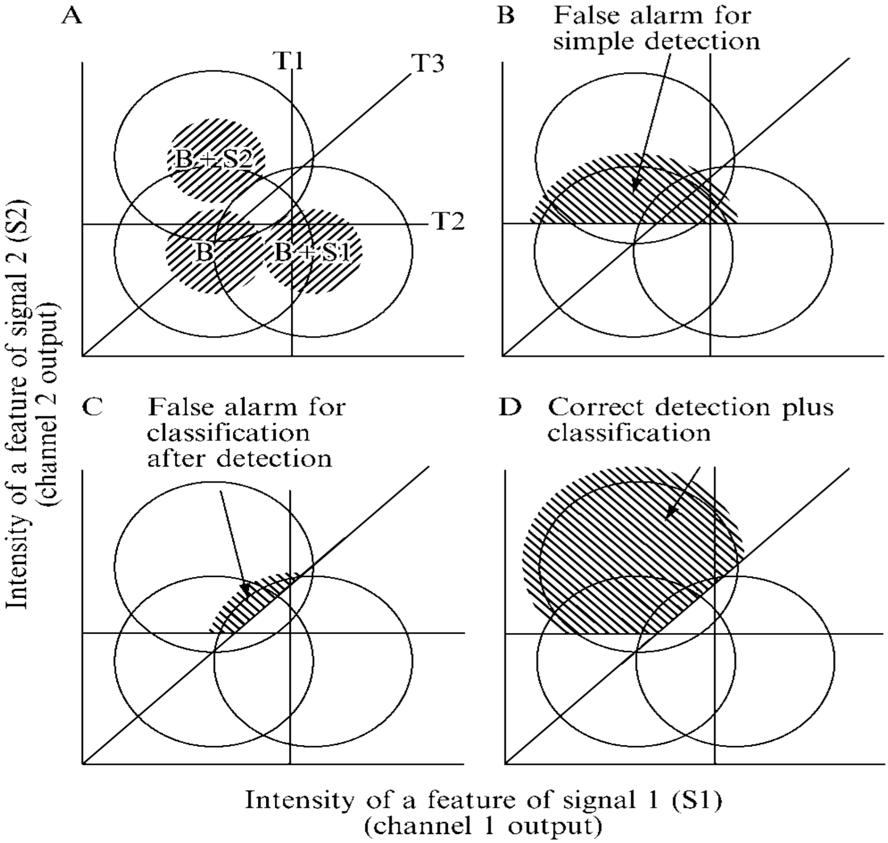


Figure 5. (A) A bivariate plot of probability densities for combined detection plus classification shows the PDFs (now represented topographically by circles of equal probability density) for background stimulation alone, B, and in combination with each of two signals, S1 and S2. Decisions in this case require three thresholds: T1 for detection of B + S1 from B; T2 for detection of B + S2 from B; and T3 for classification of a signal once detected. (B) Threshold T2 results in a P_{FA} (shaded) for responses appropriate for S2 when background alone occurs. (C) Threshold T3 results in a P_{FA} (shaded) for responses appropriate for S2 when S1 occurs. (D) A combination of thresholds T1 and T3 results in a P_{CD} (shaded) for correct detection and classification of S2.

References

Aubin, T., and Jouventin, P. (1998). Cocktail-party effect in king penguin colonies. *Proceedings of the Royal Society of London* B 265, 1665–1673.

Benzschawel, T., and Cohn, T. E. (1985). Detection and recognition of visual targets. *Journal of the Optical Society of America* A 2, 1543–1550.

Brenowitz, E. A. (1982). The active space of red-winged blackbird song. *Journal of Comparative Psychology* A 147, 511–522.

Broadbent, D. E. (1958). *Perception and Communication*. Pergamon Press, London.

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review* 74, 1–15.

Brumm, H. (2004). The impact of environmental noise on song amplitude in a territoral bird. *Journal of Animal Ecology* 73, 434–440.

Brumm, H., and Todt, D. (2002). Noise-dependent song amplitude regulation in a territorial songbird. *Animal Behaviour* 63, 891–897.

Brumm, H., Voss, K., Ko¨llmer, I., and Todt, D. (2004). Acoustic communication in noise: Regulation of call characteristics in a New World monkey. *Journal of Experimental Biology* 207, 443–448.

Cary, M., and Reder, L. M. (2003). A dual-process account of the list-length and strength- based mirror effects in recognition. *Journal of Memory and Language* 49, 231–248.

Cherry, E. C. (1953). On the recognition of speech with one, and with two ears. *Journal of the Acoustical Society of America* 25, 975.

Cherry, E. C., and Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one, and with two ears. *Journal of the Acoustical Society of America* 26, 554–559.

Cohn, T. (1978). Detection of 1-of-M orthogonal signals: Asymptotic equivalence of likelihood ratio and multiband models. *Optical Letters* 3, 22–23.

Cynx, J., Lewis, R., Tavel, B., and Tse, H. (1998). Amplitude regulation of vocalizations in noise by a songbird, *Taeniopygia guttata*. *Animal Behaviour* 56, 107–113.

Davies, D. R., and Parasuraman, R. (1982). *The Psychology of Vigilance*. Academic Press, London.

Davies, D. R., and Tune, G. S. (1970). *Human Vigilance Performance*.

Staples, London.

Dobbins, I. G., Kroll, N. E. A., and Yonelinas, A. P. (2004). Dissociating familiarity from recollection using rote rehearsal. *Memory & Cognition* 32, 932–944.

Dooling, R. J. (1982). Auditory perception in birds. In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, Eds.), Vol. 1, pp. 95–130. Academic Press, New York.

Egan, J. P., Schulman, A. I., and Greenberg, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *Journal of the Acoustical Society of America* 31, 768–773.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (1961a). Operating characteristics, signal detectability and the method of free response. *Journal of the Acoustical Society of America* 33, 993–1007.

Egan, J. P., Greenberg, G. Z., and Schulman, A. I. (1961b). Interval of time uncertainty in auditory detection. *Journal of the Acoustical Society of America* 33, 771–778.

Endler, J. A. (1990). On the measurement and classification of color in studies of animal colour patterns. *Biological Journal of the Linnean Society London* 41, 315–352.

Endler, J. A. (1993). The color of light in forests and its implications. *Ecological Monographs* 63, 1–27.

Endler, J. A., and Théry, M. (1996). Interacting effects of lek placement, display behavior, ambient light, and color patterns in three Neotropical forest-dwelling birds. *American Naturalist* 148, 421–452.

Enquist, M., and Arak, A. (1993). Selection of exaggerated male traits by female aesthetic senses. *Nature* 361, 446–448.

Glanzer, M., and Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*13, 8–20.

Glanzer, M., Adams, J. K., Iverson, G. J., and Kim, K. (1993). The regularities of recognition memory. *Psychological Review* 99, 546–567.

Gomez, D., and Théry, M. (2004). Influence of ambient light on the evolution of colour signals: Comparative analysis of a Neotropical rainforest bird community. *Ecology Letters* 7, 279.

Green, D. M. (1961). Detection of auditory sinusoids of uncertain frequency. *Journal of the Acoustical Society of America* 33, 897–903.

Green, D. M., and Birdsall, T. G. (1978). Detection and recognition. Psychol. Rev. 85, 192–206. Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. Wiley & Sons, New York.

Green, D. M., Weber, D. L., and Duncan, J. E. (1977). Detection and recognition of pure tones in noise. *Journal of the Acoustical Society of*

America 62, 948–954.

Guilford, T., and Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour* 42, 1–14.

Guilford, T., and Dawkins, M. S. (1993). Receiver psychology and the design of animal signals. *Trends in Neurosciences* 16, 430–436.

Heindl, M., and Winkler, H. (2003). Vertical lek placement of forest-dwelling manakin species (Aves, Pipridae) is associated with vertical gradients of ambient light. *Biological Journal of the Linnean Society London* 80, 647–658.

Klein, S. A. (1985). Double-judgment psychophysics, problems and solutions. *Journal of the Optical Society of America* A 2, 1560–1585.

Klump, G. M. (1996). Bird communication in the noisy world. In *Ecology and Evolution of Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, Eds.), pp. 321–338. Cornell University Press, Ithaca.

Klump, G. M., and Langemann, U. (1995). Comodulation masking release in a songbird. *Hearing Research* 87, 157–164.

Langemann, U., and Klump, G. M. (2001). Signal detection in amplitude-modulated maskers: I. Behavioral auditory threshold in a songbird. *European Journal of Neuroscience* 13, 1025–1032.

Langemann, U., Gauger, B., and Klump, G. M. (1998). Auditory sensitivity in the great tit perception of signals in the presence and absence of noise. *Animal Behaviour* 56, 763–769.

Leonard, M. L., and Horn, A. G. (2005). Ambient noise and the design of begging signals. *Proceedings of the Royal Society of London Biological Sciences* 272, 651–656.

Loeb, M., and Alluisi, E. A. (1977). An update of findings regarding vigilance and a reconsideration of underlying mechanisms. In *Vigilance: Theory, Operational Performance, and Physiological Correlates* (R. R. Mackie, Ed.), pp. 719–749. Plenum Press, New York.

Luce, R. D. (1963). A threshold theory for simple detection experiments. *Psychological Review* 70, 61–79.

Luce, R. D., and Green, D. M. (1974). Detection, discrimination and recognition. In *Handbook of Perception* (E. C. Carterette and M. P. Friedman, Eds.), Vol. 2. Academic Press, New York.

Lynn, S. K., Cnaani, J., and Papaj, D. R. (2005). Peak shift discrimination learning as a mechanism of signal evolution. *Evolution* 59, 1300–1305.

Mackie, R. R. (Ed.) (1977). *Vigilance: Theory, Operational Performance, and Physiological Correlates*. Plenum Press, New York.

Macmillan, N. A. (2002). Signal detection theory. In *Stevens' Handbook of Experimental Psychology, Third Edition. Methodology in Experimental Psychology* (H. Pashler and J. Wixted, Eds.), Vol. 4, pp. 43–90. John Wiley & Sons, New York.

Macmillan, N. A., and Creelman, C. D. (1990). Response bias, characteristics of detection theory, threshold theory, and "nonparameteric" indexes. *Psychological Bulletin* 107, 401–413.

Macmillan, N. A., and Creelman, C. D. (1991). '*Detection Theory: A User's Guide*. Cambridge University Press, Cambridge.

Macmillan, N. A., and Kaplan, H. L. (1985). Detection theory analysis of group data, estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin* 98, 185–199.

Mulligan, R., and Shaw, M. L. (1980). Multimodal signal detection, independent decisions vs. integration. *Perception & Psychophysics* 28, 471–478.

McNicol, D. (1972). '*A Primer of Signal Detection Theory*. Allen and Unwin, London.

Naguib, M. (2003). Reverberation of rapid and slow trills: Implications for signal adaptations to long range communication. *Journal of the Acoustical Society of America* 133, 1749–1756.

Naguib, M., and Wiley, R. H. (2001). Estimating the distance to a source of sound: Mechanisms and adaptations for long-range communication. *Animal Behaviour* 62, 825–837.

Nieder, A., and Klump, G. M. (2001). Signal detection in amplitude-modulated maskers: II. Processing in the songbird's auditory forebrain. *European Journal of Neuroscience* 13, 1033–1044.

Nolte, L. W., and Jaarsma, D. (1967). More on the detection of one of M orthogonal signals. *Journal of the Acoustical Society of America* 41, 497–505.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America* A 2, 1508–1532.

Pierce, J. (1963). Some sources of artifact in studies of tachistoscopic perception of words. *Journal of Experimental Psychology* 66, 363–370.

Pollack, I., Rubenstein, H., and Decker, L. (1959). Intelligibility of known and unknown message sets. *Journal of the Acoustical Society of America* 31, 273–279.

Richards, D. G. (1981a). Alerting and message components in songs of rufous-sided towhees. *Behaviour* 76, 223–249.

Richards, D. G. (1981b). Estimation of distance of singing conspecifics by the Carolina wren. *Auk* 98, 127–133.

Seehausen, O., van Alphen, J. J. M., and Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science* 277, 1808–1811.

Shaw, M. L. (1982). Attending to multiple sources of information, I. The integration of information in decision making. *Cognitive Psychology* 14, 353–409.

Slabbekoorn, H., and Peet, M. (2003). Birds sing at a higher pitch in urban noise. *Nature (London)* 424, 267.

Starr, S. J., Metz, C. E., Lusted, L. B., and Goodenough, D. J. (1975). Visual detection and localization of radiographic images. *Radiology* 116, 533–538.

Swensson, R. G., and Judy, P. F. (1981). Detection of noisy visual targets: Models for the effects of spatial uncertainty and signal-to-noise ratio. *Perception & Psychophysics* 29, 521–534.

Swets, J. A., and Birdsall, T. G. (1978). Repeated observation of an uncertain signal. *Perception & Psychophysics* 23, 269–274.

Swets, J. A., Shipley, E. F., JMcKee, J. M., and Green, D. M. (1959). Multiple observations of signals in noise. *Journal of the Acoustical Society of America* 31, 514–521.

Tanner, W., and Swets, J. (1954). The human use of information—I: Signal detection for the case of the signal known exactly. *IEEE Transactions on Information Theory* 4, 213–221.

Tanner, W. P., Jr. (1956). Theory of recognition. *Journal of the Acoustical Society of America* 28, 882–888.

ten Cate, C., and Bateson, P. P. G. (1988). Sexual selection: The evolution of conspicuous characteristics in birds by means of imprinting. *Evolution* 42, 1355–1358.

Thomas, J. P. (1985). Detection and identification, how are they related? *Journal of the Optical Society of America* A 2, 1457–1467.

Uy, J. A. C., and Endler, J. A. (2004). Modification of the visual background increases the conspicuousness of Golden-collared Manakin displays. *Behavioral Ecology* 15, 1003–1010.

Warm, J. S. (1977). Psychological processes in sustained attention. In '*Vigilance: Theory, Operational Performance, and Physiological Correlates* (R. R. Mackie, Ed.), pp. 623–644. Plenum Press, New York.

Watson, C. S., and Nichols, T. L. (1976). Detectability of auditory signals presented without defined observation intervals. *Journal of the Acoustical Society of America* 59, 655–668.

Weary, D., Guilford, T. C., and Weisman, R. G. (1993). A product of discrimimay lead to female preferences for elaborate males. *Evolution* 47, 333–336.

Wiley, R. H. (1983). The evolution of communication, information and manipulation. In '*Animal Behavior, Volume 2, Communication* (T. R. Halliday and P. J. B. Slater, Eds.), pp. 131–181. Blackwell Scientific Publications, Oxford.

Wiley, R. H. (1991). Associations of song properties with habitats for territorial oscine birds of eastern North America. *American Naturalist* 138, 973–993.

Wiley, R. H. (1994). Errors, exaggeration, and deception in animal communication. In '*Behavioral Mechanisms in Evolutionary Ecology* (L. R. Real, Ed.), pp. 157–189. University of Chicago Press, Chicago.

Wiley, R. H., and Richards, D. G. (1982). Adaptations for acoustic communication in birds, sound propagation and signal detection. In *Acoustic Communication in Birds* (D. E. Kroodsma and E. H. Miller, Eds.), Vol. 1, pp. 131–181. Academic Press, New York.

Wollerman, L., and Wiley, R. H. (2002). Background noise from a natural chorus alters female discrimination of male calls in a Neotropical frog. *Animal Behaviour* 63, 15–22.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language* 46, 441–517.

Chapter 9

Communication as a Transfer of Information

Introduction

No one seems ever to have doubted that animals can communicate with each other. The evidence for communication has always seemed obvious – responses by one individual to the actions of another. In his extended discussions of animal behavior, Darwin for instance took communication by animals for granted. Although he cited many reports of animals' responses to each other, he never made them the subject of his studies. Instead, he focused on evidence for continuity between humans and non-human animals in the evolution of mental processes. In *The Descent of Man, and Selection in Relation to Sex* (1871), he made an extended case that animals express many of the same emotions that humans do, even such mental activities as deceit, revenge, humour, deliberation and reason. In *The Expression of the Emotions in Man and Animals* (1873), he elaborated on phylogenetic continuity in the expression of many emotions – although not all of those he had mentioned previously. Furthermore, his principle of antithesis, that contrasting emotions tended to be associated with contrasting actions, suggested that animals' actions evolved by natural selection to promote communication.

The basic components of communication are now widely recognized – signaler, signal, and receiver. To confirm that communication has occurred, it is thus necessary to show that one individual has produced a signal – a pattern of stimulation – to which another individual has responded. Experimental investigation of this process began with the use of simple models by early ethologists such as Niko Tinbergen (1951). In recent decades, presentations of audio and video recordings and even robotic models have resulted in extensive experimental analysis of communication by animals.

Yet Darwin's principal claim remains controversial. Is there continuity between mental processes of humans and those of other animals? Even if the differences prove to be qualitative, can we measure the magnitude of the differences? As Darwin recognized, one of the central issues in these controversies is communication. What do animals communicate? And how much do they communicate? These questions are often phrased in terms of information. What information is communicated by animals? And how much?

This chapter addresses these questions in four steps. First, it reviews the concept of information in communication and thereby concludes that all communication must involve a transmission of information. Second, it considers, but rejects, the argument that information and manipulation are incompatible. Third, it argues that the transfer of information depends on mental processes of categorization and association. Fourth, it addresses the issue of information about mental states of other individuals and ourselves. It concludes with an element of necessary ignorance.

Communication as a Transfer of Information

It is probably not a coincidence that the three components of communication – signaler, receiver and signal – were first identified in the decades following the invention and deployment of the telegraph. De Saussure's (1916 [1959]) diagram takes the telegraph as a metaphor for human language, and Ogden and Richards (1923) elaborate the model by emphasizing the mental processes of the signaler and receiver. Linguists and philosophers now use these models routinely in their discussions of communication.

Further advances in engineering and the widespread adoption of telephones and electromagnetic radiation for human communication eventually led to competition for communication. How many radio stations could simultaneously operate in one area? How many conversations could simultaneously use one telephone line? Investigation of these practical issues revealed that communication had limits. To understand these limits, it was apparent that communication had to be measured. Shannon's (1948, 1963) pioneering contribution was to propose a measure of information and then to use it to demonstrate mathematically that the properties of the connection between signaler and receiver – the channel – imposed a limit on the amount of information that could be transmitted in any period of time.

Shannon's measure of information in a set of *i* signals (H_0) equals $\sum p_i \ln p_i$, with p_i being the probability of the *i*th signal. As Shannon explains, this particular expression is the simplest one possible that can satisfy our intuitive requirements for a measure of the amount of communication. This measure (H_0) is the number of binary decisions required to specify which signal in a message is next, or in other words to specify the occurrence of any one signal. It is thus the uncertainty in predicting the occurrence of any one signal. An informative message would have high uncertainty about the occurrence of any one signal (it would require many binary decisions to specify each signal's occurrence). Frequent use of just a few signals conveys less information than would many less frequent signals. An infrequent signal increases the amount of information in a message more than does a frequent signal.

A set of signals could consist of a sequence of signals in time or an arrangement of signals in space. Shannon's measure applies to both cases. In either case, identifying a set of signals often requires some method for segmenting the temporal or spatial continuity of an animal's actions into components. As Shannon shows, this segmentation is not necessary, because his conclusions still apply in the limit of continuously varying signals and responses. Nevertheless most attempts to measure information require segmentation of animals' actions into sets of signals and responses.

The concept of information as a measure of the degree of uncertainty in a pattern of signals contrasts with the usual concept of information as the degree of certainty a receiver acquires from signals. Shannon's definition of information thus seems contrary to any definition that others might accept as intuitively appropriate. The issue is whether information is a property of the structure of signals or of the state of the receiver.

The problems arising from the segmentation of actions and the nature of information have both resulted in distortions in how biologists think about information. Both have resulted in premature rejections of information in animal communication. The remainder of this section discusses the first of these issues, the segmentation of signals and responses. Subsequent sections take up the second issue, the receiver's state of mind.

The problem of segmentation arose during attempts to measure the amount of information in animals' displays. At the time it seemed that such measures would allow comparisons of communication by different species or different modalities. Attempts to measure the information in the displays of rhesus macaques and fiddler crabs (Altmann, 1965; Hazlett & Bossert, 1965) consisted of identifying a set of distinct actions and then estimating the probability of each in particular contexts. Ethologists had become accustomed to describing ethograms, discrete categories of actions for each species. Measuring the information in displays was just one of a number of ways that ethograms could be used to quantify behavior. It became apparent, however, that any measure of information depended on how the observer chose to segment the animals' actions. In some cases, such as the songs and stereotyped displays of some birds, actions seem relatively invariant and discrete, although only in a few cases has variation actually been measured (Wiley, 1973). As a rule, however, animals' displays, including those of monkeys and crabs, consist of variable and intergrading actions. When an observer segments these variable displays into discrete categories, the number of categories and their boundaries have unknown relevance for the animals involved. As a result, measuring the amount of information in animal displays seems arbitrary, and comparisons of different species seem fruitless. Only a few studies have followed these precedents (Dingle, 1969; Steinberg & Conant, 1974).

The problem of segmentation is not insurmountable, however. The problem lies not with segmentation of animals' actions in itself but with identifying a segmentation relevant to the species under investigation, rather than one imposed by a human observer. Behavioral and neurophysiological experiments can determine how individuals classify stimulation. Yet we often do not know as much as we should.

Consider recognition of conspecific individuals. Many experiments have shown that animals respond to signals from their own species but not from others, and that they respond to signals from particular individuals but not others (Falls, 1982; Wiley & Wiley, 1977; Godard & Wiley, 1995). Recognition of young by parental birds provides a good example. Beecher and his students, for instance, have shown that adult cliff swallows recognize their own young while they are still in the nest, whereas adult barn swallows do not (Beecher et al., 1986). When nestling barn swallows were experimentally exchanged between nests, parents responded to nestlings from another nest just as much as to those of their own. In contrast, parent cliff swallows did not feed others' young under any conditions. Playbacks of nestlings' calls confirmed this difference between the two species in parental recognition of young. Furthermore, the calls of nestling cliff swallows vary more among individuals than those of barn swallows, as expected if they encode more information about individual identity. Because the nests of cliff swallows are clustered in dense colonies, while those of barn swallows are dispersed, only among cliff swallows might parents or young occasionally enter the wrong nest. Since cliff swallows build flask-shaped nests of mud, so the young inside are in nearly complete darkness, it makes sense that the nestlings' vocalizations have evolved to promote parental recognition.

Although parent cliff swallows recognise their own young at least collectively, it is not known whether or not they can go a step farther to recognise each of their young individually. Individual recognition of this sort would require that parents respond to each offspring in a distinctive way. Current experiments indicate only that parents distinguish familiar nestlings' calls from unfamiliar ones.

These classic experiments demonstrate two important points. First, animal communication does convey information, in this case the identity of offspring, and, moreover, the analogous signals of two different species can differ in the amount of information conveyed. Second, even in species for which transmission of information has been demonstrated, it is a more difficult task to determine how much information this is. In particular, we do not know whether parents recognize their young collectively or each one individually. The difference is between a binary discrimination (between categories of their own young and all others) and a more complex discrimination (between as many as six individual young). In this case the units of classification (individual organisms) are clear, and we understand something about how swallows of different species classify these units, but there remain open questions about the complexity of this classification.

Even when units of classification are apparent, actual signals and responses themselves are likely to vary continuously or at least in complex ways. This complexity makes it difficult to measure the amount of information in signals. Beecher and colleagues (Beecher et al., 1989; Medvin et al., 1993) have estimated the potential amount of information in the vocalizations of nestling barn and cliff swallows by measuring variation in many different features of frequency and timing and then reducing this variation to a set of independent principal components. They could then use the standard deviations of these principal components to estimate the potential amount of information, in binary units, that these vocalizations contain. This estimate is an upper limit for the amount of information transmitted from signaler to receiver. To determine the actual amount of information transferred would require experiments to document the association between variation in signals and variation in responses.

Haldane and Spurway (1954) had earlier used similar procedures to determine the amount of information transmitted by the waggle dances of honeybees. Variation in the directions of honeybees' foraging flights provided an estimate of the amount of directional information that foraging bees obtained from waggle dances. Error in the mean direction of foraging flights provided an estimate of the amount of information in the dances themselves. Haldane and Spurway concluded that the dances appeared to contain two to three bits of information more than the accompanying foragers received. Recent measurements have shown that variation in the directions of the dances themselves depends on the distance or nature (food or nest site) of the target (Towne & Gould, 1988; Weidenmu¨ ller & Seeley, 1999; Tanner & Visscher, 2010). Similar procedures have been used to estimate the amount of information in the odor trails of ants (Wilson, 1962).

These cases show that determining the amount of information in animal signals must clear some technical hurdles (Pfeifer, 2006), but they also indicate that this process is important if we wish to understand the complexity of animal communication. As discussed below, understanding the categorization of stimulation by animals is critical for any understanding of the evolution of communication.

Manipulation versus Information

Prior to the development of rigorous thinking about the evolution of cooperation, it had always seemed that communication was an example of cooperation. Signalers provided information that receivers used. In The Behavior of Communicating (1980), Smith took this point for granted. He identified the 'message' of a signal as the association between a signaler's action and its current state (its neural and physiological state, including its disposition to act in particular ways and its perception of its environment). Thus the message of a signal might be that the signaler is likely to fight if attacked, that it has just seen a predator approaching, that it is in excellent physical condition, or that it is a particular species or individual. Marler (1961) had earlier discussed the ways in which signals are associated with states of signalers. Smith then identified the 'meaning' of a signal as the association between the signal and the receiver's responses. He made the important point that signals with the same message could have different meanings for receivers, as a result of differences in a receiver's context and state. Much of Smith's own work focused on determining the associations between the signals produced by birds and their contexts and states.

Although these early analyses assumed that communication had mutual advantages for signaler and receiver, they nevertheless emphasized two undeniable features of communication. Signals include information only by virtue of their associations with the states (including contexts) of signalers. They transmit information only by virtue of their associations with the responses of receivers.

Rejecting the assumptions that communication is necessarily mutually beneficial to the participants, Dawkins and Krebs (1978; Krebs & Dawkins, 1984) argued that signals instead evolve by natural selection to manipulate receivers to respond in ways that provide advantages for the signaler, regardless of any advantages for the receiver. This position, however, raised the question of why receivers should respond to signals in ways that were disadvantageous for them. A possible answer is that signalers exploit sensory biases of receivers, in other words constraints on the way receivers respond to signals (Guilford & Dawkins, 1991; Arak & Enquist, 1993; Endler & Basolo, 1998; Ryan, 1998). Such constraints might occur when receivers have evolved to respond to particular signals in another context. For instance, if females respond to particular signals in finding food or shelter, a signal mimicking these signals might stimulate a female to respond to a male when she otherwise would not. Alternatively receivers might respond to exaggerated signals not normally produced by signalers, examples of supernormal stimuli, as a result of the retention of ancestral constraints on their nervous systems or as a result of peak shift in learning (Hogan, Kruijt & Frijlink, 1975; Lynn, Cnaani & Papaj, 2005; ten Cate & Rowe, 2007). Both of these proposals assume that receivers have not yet evolved more discriminating responses to sensory input.

A revision of this position came when Grafen (1990) emphasized that receivers must on average receive benefits from their responses, otherwise natural selection would tend to eliminate those responses. As a rule receivers should respond only to signals that convey information about (are associated with some feature of) the signaler that is useful to the receiver. Grafen attempted to confirm Zahavi's (1977; restated in a 1997 book) proposal that the cost of a signal guarantees its honesty, because for one reason or another the cost is too great for a deceptive signaler to bear. Searcy and Nowicki (2005), in their review of the evolution of communication, confirmed the three relevant points in this theory of honest signaling: (1) to qualify as signals, actions must at least occasionally evoke responses from receivers; (2) receivers must on average benefit from their responses; and (3) signals must convey information about signalers in ways related to their costs. For instance, females respond to signals associated with high-quality mates, and these signals have higher costs than those associated with low-quality mates. Searcy and Nowicki identify many examples of animal communication that meet these criteria.

These conditions for honest communication are close to the position that Smith advocated earlier, with the addition of a stipulation that signals are honest on average. The message of a signal is its association with the state of the signaler. Receivers on average benefit from their responses. Communication is on average honest and thus normally advantageous for both signaler and receiver. This revised position leaves open the possibility for manipulation, which occurs when signalers can take advantage of receivers by mimicking a signal that would in other contexts evoke a response beneficial to the receiver or when receivers can take advantage of signalers by eavesdropping on signals that would in other contexts evoke responses beneficial to the signaler. In all such cases receivers and signalers benefit on average from communication, although on infrequent occasions they are manipulated to their disadvantage.

Numerous such cases of manipulation are now known. For instance, birds occasionally give false alarms for predators in order to gain access to food that is otherwise monopolized by more dominant individuals (Møller, 1988), and some primates practice deception routinely (Whiten & Byrne, 1988; Cheney & Seyfarth, 1991; Mitchell & Anderson, 1997). It is now apparent that signals must, as a result of natural selection, evoke responses that have advantages for both signaler and receiver, at least on average. Manipulation is thus the exception that proves the rule (Wiley, 1994).

Recent discussions of the role of information in animal communication emphasize one aspect of these conclusions or another but are not actually in conflict (Owings & Morton, 1998; Rendall et al., 2009; Seyfarth et al., 2010). Signals do convey information about the signaler, and yet sometimes signals are manipulative. Everybody can agree that communication has three basic components: signals include information (about the signaler or its situation), receivers respond (overtly or covertly, with high or low probability), and both signalers and receivers benefit on average.

These conclusions do not resolve all problems raised by information in signals. Most people feel that the information in signals is more than a correlation with the internal or external state of the signaler. Instead, most people feel that information is about something – about something the signaler perceives or thinks. De Saussure (1916 [1959]) emphasized the relationship between a signal and the signaler's mind, and Ogden and Richards (1923) added an external referent to make a triangular relationship – signal, mind and referent – necessary to understand the 'meaning of meaning'. In the following sections, I pursue an engineering approach to information in order to address the 'meaning of meaning'.

Communication as Categorization and Association

So far I have used the term 'signal' loosely. Engineers in fact never seem much concerned with a definition of a signal, although ethologists have perennially wrestled with a definition. Maynard Smith and Harper (2003), like Grafen (1990) previously, emphasized that a signal must have evolved for the purpose of evoking a response. They maintain that the condition of evolution distinguishes signals from 'cues', which include inanimate sources of stimulation that influence animals' behavior. They also distinguish two kinds of signals – indices and handicaps – based on whether or not a signal is reliable because it cannot be faked or has excessive costs. All signals have costs, but handicaps have 'strategic costs', costs in excess of 'efficacy costs' which are those 'needed to transmit the information unambiguously' (Maynard Smith & Harper, 2003, p. 7).

Shannon and Weaver (1963) defined ambiguity as the uncertainty in responses to a given signal (as opposed to equivocation, uncertainty in signals for a given response). Ambiguity and equivocation are the two components of noise in any system of communication. Ambiguity is a relationship between a signal and responses, not a property of a signal. In the real world, as Shannon realized, there is no communication without noise – no communication without ambiguity. Communication can have more or less ambiguity (and equivocation), more or less efficacy, if you will, but there is no transition between efficacious and 'strategic' communication, as Maynard Smith and Harper maintain. All the costs of an evolved signal contribute to efficacy, to reducing ambiguity by evoking an appropriate response.

Instead I have proposed a definition of a signal that does not require an antecedent understanding of its evolution (Wiley, 1994). A signal is any pattern of energy or matter that evokes a response without providing all of the power for that response. For instance, if a tree falls, shoving someone out of the way is not communication (the shove is not a signal because it provides sufficient energy to effect the movement of the recipient). A shout, "Heads up!", on the other hand, is a signal, provided the receiver sometimes responds, for instance by jumping out of the way. By this definition, the sound of the cracking trunk is also a signal, so that, if the hearer jumps away, communication has also occurred.

Two points need clarification. First, a system of communication includes many such instances of signals, not just one. It is characterized by probabilities, not isolated instances. Communication occurs even if on some occasions the recipient does not move. Second, although this definition includes inanimate sources of signals (which Maynard Smith and Harper and others would term cues), there is no essential distinction between these and signals from animate sources. However, animate sources can evolve, which raises the possibility that the properties of signals become optimized for communication, as Darwin's principle of antithesis had suggested. Because a signal does not provide all of the power for a response, the receiver must perceive a signal and provide some, often most, of the power for the response. The signal of course must provide enough power to affect the receiver's sensory organs.

A receiver thus must have a nervous system (or some other feature of its physiology), which perceives a signal and then associates it with a response, and a musculo-skeletal system (or in some organisms just a muscular system), which produces the response. The engineering equivalents are transducers, gates (switches) and amplifiers. The definition of a signal thus also defines a receiver – a mechanism with transducer, gate and amplifier.

Second, this definition makes it clear that the receiver is in control of communication. A receiver is a decision-making mechanism. It categorizes impinging stimulation into perceived signals and associates perceived signals with responses. Animate receivers can evolve and thus optimize responses to a signal. As I have discussed elsewhere (Wiley, 1994, 2006) and as Grafen (1990) had previously recognised, such receivers should in all cases benefit on average from their responses to signals. They should evolve to minimise responses to unreliable or deceptive signals. In communication there is always the possibility of some deception or error, but unless responses to signals provide some benefit on average to a receiver, receivers should evolve to cease responding altogether.

Instead of pursuing questions about the evolution of communication, I want to emphasize here the decision-making property of receivers. One of their essential properties is categorizing the stream of incoming stimulation. This is the process often called segmentation, by which discrete objects or units are identified in the continuous flood of stimulation. It is the basis of perception – a relationship between the properties of stimulation and the properties of the sensory components of the receiver's nervous system (the sense organs and higher sensory centers of the nervous system). The examples of recognition of conspecifics above are clear cases of the categorization of sensory input. In fact, all communication involves such categorization.

Furthermore, a receiver must associate perceived categories of stimulation (signals) with responses. As Sherrington (1906) long ago emphasized, one important function of all nervous systems is to associate each perceived category of stimulation with a particular response or suite of responses. The motor components of the nervous system, which control the musculoskeletal system, become the 'final common pathway'. Categorization of stimulation and association of the resulting categories with responses are the two fundamental properties of an animate receiver – and indeed of any nervous system.

The process of categorizing stimulation raises a question about the perceptual demarcations of categories. In general terms the rules for demarcation could be learned or innate. By innate, I mean developing in the absence of sufficient external information to specify the resulting rule. For instance, the striate cortex of mammals includes cells that develop into stripe detectors that respond only to strips of light in particular orientations at particular locations in the animal's visual field. It is now well known that these cells develop before birth in the absence of any exposure to patterns of light such as stripes (Hubel & Wiesel, 1963). Their development requires only environmental conditions sufficient for normal development of the brain in general. After the eyes open, the further development of these cells depends on subsequent environmental conditions, both general (exposure to light regardless of pattern) and specific (exposure to particular patterns of light, such as predominantly horizontal or predominantly vertical stripes). Nevertheless, under normal conditions for brain development, these cells develop initially to detect specific features without exposure to patterns of light.

Songbirds learn features of their songs, but they begin the process of learning with innate predispositions to attend to certain patterns of sound (in some cases, components of conspecific songs) or to sounds in certain situations (Marler, 1984; Marler & Peters, 1988). Without these initial (innate) predispositions, it is hard to imagine how a naive bird could identify in the flood of incoming stimulation what it is that it should learn.

The importance of predispositions for learning had previously been emphasized by Lorenz (1966). Chomsky (1959, 1986) made the same point about the development of language in human children. Much earlier, Kant (1793 [1961]) had elaborated his philosophy from similar arguments for innate categories in all rational thought. Association without predisposition leads to chaos. With predispositions, association can produce the extraordinary complexity of animal and human behavior, much if not most of which is communication in one form or another.

The expanded brain of humans no doubt allows greater complexity in categorization and association. Quantitative increase in components can, in some sense, lead to qualitative differences in performance. Of course, recognizing a qualitative difference is itself a form of categorization. Humans are perhaps predisposed to recognize categories of human and non-human animals. So caution should temper any conclusion that human and non-human brains differ categorically.

Communication of States of Mind

If the nervous systems of organisms, their brains in particular, are mechanisms for categorizing stimulation and for associating the resulting categories with responses, are they sufficient to produce minds? Whether brains are sufficient explanations of minds is, of course, an old question in philosophy – perhaps, in one form or another, the only question. If an organism's mind is sufficiently explained by its brain, then the minds of other organisms are revealed by their behavior. This position, I suggest, is the essence of behaviorism (Morris, 1955; Bennett, 1976) (perhaps philosophical behaviorism is a better label, to distinguish it from psychological behaviorism). What a stimulus means to me, for instance, is entirely equivalent to how it changes my behavior. The change might not be immediate or overt. In addition, because so much of behavior is communication in one form or another, we should accept probabilistic changes in behavior as meaning. Two signals would have different meanings if they evoked different probability distributions of responses, covert or overt, immediate or eventual.

With this point of view, a 'theory of mind' would consist of an ability to predict, at least probabilistically, other individuals' responses to stimulation (signals). Humans obviously can manage this feat. As was long ago pointed out to me, we would not dare to drive home if we could not predict other people's behavior. It is also obvious that all animals can predict the behavior of other individuals, at least probabilistically, and respond appropriately. This behavioral point of view thus implies that all animals have 'theories of mind'. Just as animals differ in complexity of associative learning, so their 'theories of mind' differ in complexity.

Mind, however, is not obviously equivalent to brain. This nagging reservation arises particularly when I consider my own mind. Sometime in the future neurophysiologists might be able to specify the precise state of every cell in my brain at a particular time. It might, for instance, become possible to specify exactly which neurons are activated when I see a particular tree or when I imagine the concept of *treeness*. Nevertheless, these neurons, it is easy to suppose, might not be my image of a tree nor my concept of a tree. The issue is whether observations of another organism's brain or behavior are, or are not, enough to characterize its mind. Such thoughts raise many issues, but at the core of these issues are self-awareness and intention.

Evidence for self-awareness, it is often assumed, can come from reactions to mirrors. When an animal or human directs responses to its own body while viewing itself in a mirror, it seems that it must have an awareness of itself (Gallup, 1970, 2011; De Veer & Van den Bos, 1999). This ability, however, develops as a result of experience with mirrors. Humans without such experience are baffled by mirrors. Indeed, even photographs and recordings confuse people who lack experience of seeing or hearing themselves. Learning that visual images in a mirror can be associated with actions directed toward one's own body requires mastery of a chain of contingencies. It requires a complexity of learning that is evidently beyond the abilities of most animals. Yet it is not clear that the process differs qualitatively from other examples of associative learning. Furthermore, anyone who has tried using mirrors to view the back of the head, or, worse still, to direct actions there, becomes quickly disabused of any conclusion that self-awareness is equivalent to mastery of mirrors.

Normally our sense of self-awareness comes from introspection, just as does our sense of *treeness* or a particular tree. It is probable that association is important for this introspective self-awareness. We might associate all of our responses to sensory input with a common agent, in other words, our self. Once again, it might become possible, sometime in the future, to predict when a person is self-aware by determining the state of neurons in the brain, but making this connection would depend on the person's own report of self-awareness. Thus it is not clear that a description of neurons can ever be equivalent to self-awareness. Even one-to-one mapping of behavior and brain might not guarantee existential equivalence of mind and brain.

I am trying to choose my words carefully here so as not to take a position that mind is, or is not, brain. The preceding arguments suggest that currently there is no indisputable evidence for or against either position. At least to my mind, if I may say so, it is not utterly obvious that mind is brain, nor that it is not. Nevertheless, a discussion of information in communication must include attention to the state of mind often thought to be crucial for human communication – intention.

Intentions (in the sense of preconceptions of actions) are often assumed to distinguish human communication from that of animals. When communicating, I intend to modify the recipient's mind, at least in some minimal way. When speaking of a tree, for instance, I intend for the listener to acquire an image of a tree somewhat like my own. As Wittgenstein (1968) has famously emphasized, this process requires that we have developed similar rules for using signals, or, Shannon (1948) would say, for encoding and decoding signals. We must both associate the word 'tree' with a mental image such as 'generalized tree'. These rules are just as important when our intention is to deceive. Intentions are the basis for much human morality and justice. It has also been proposed that adopting an 'intentional stance' (Dennett, 1987) facilitates communication or indeed all interactions with animate and perhaps even inanimate objects. What produces intentions? And how do we recognize them in others?

By introspection, my intentions seem connected to my awareness of my self as an agent. As described previously, this awareness might arise as a result of an association of my responses. It is thus a second-order association. Particular sensory inputs become associated with certain responses. Then these sensory– motor associations become associated with each other to produce a sense of agency. It is the patterns in my behavior as a sentient and responsive organism that generate my sense of my self as an agent (so it appears to me on introspection). Can someone else study my intentions by studying my brain and behavior? It is not clear that they can, because, just as with self-awareness, verification of my intentions requires my introspection.

If my intentions are a result of introspection, to pursue this argument, my willingness to attribute them to others must depend on empathy. I can of course, with enough study, predict the behavior of others, in a probabilistic way, as discussed above. Yet, if mind is not behavior, attributing mental states such as self-awareness and intentions can only occur by empathy. Empathy is attribution of mental states based on a sense of similarity between oneself and another. The more similar another person is to me, for instance in behavior, the easier it is to empathize. Empathy can be extended to non-human animals, on the same basis, and even to plants and inanimate objects. To the extent that my automobile responds predictably to my input and my careful attention, I can empathize with it. It might well help me to communicate with my automobile to take an intentional stance and to empathize with it. The personification of many objects and features of the environment by peoples of many cultures could well have the same basis.

The two contrasting possibilities, that mind is brain or that it is not, thus lead to two contrasting views of information in communication. The first possibility leads to probabilistic predictions of behavior, based on associations of contexts with responses. The second leads to introspection and empathy. If attributing states of mind to other organisms is equivalent to predicting their behavior, then I regard all animals (as well as people) as having states of mind, and all as having theories of mind that allow them to respond to other individuals in appropriate ways. If attributing states of mind is not equivalent to predicting behavior, then animals (and indeed other people or even machines) have states of mind depending on my ability to empathize with them.

I want to stress once again that I take no position on this polarity. It is not clear to me whether it will be possible, sometime in the future, to reduce my introspective sense of self-awareness, intention and meaning to the states of the neurons in my brain or to complexities in my behavior. Thus it is also not clear to me that I can identify these states of mind in other individuals by studying their brains or behavior. It is an uncertainty I can live with, however. I conditionally accept that mind is brain and proceed to analyze how animals communicate, how they categorize other individuals and their environment, how they associate sensory input with responses, how complicated these processes can be. Attributing states of mind, self-awareness and intentions to other humans is a necessary feature of our moral and legal systems. Attributing these states to other people, to non-human organisms and to inanimate objects is often an amusing diversion and might also help me to interact with them fruitfully. Insofar as I do anything more than predict or anticipate their behavior, however, I engage in empathy.

To summarize, this discussion has led to three conclusions:

(1) Communication consists of transmission of information from one individual to another;

(2) If mind is behavior, then *all* organisms communicate states of mind;

(3) If it is not, then *no* communication transmits states of mind.

References

Altmann, S. A. (1965). Sociobiology of rhesus monkeys. II: Stochastics of social communication. *Journal of Theoretical Biology*, 8, 490–522.

Arak, A. & Enquist, M. (1993). Hidden preferences and the evolution of signals. *Philosophical Transactions: Biological Sciences*, 340, 207–213.

Beecher, M. D. (1989). Signaling systems for individual recognition: an information theory approach. *Animal Behaviour*, 38, 248–261.

Beecher, M. D., Medvin, M. B., Stoddard, P. K. & Loeschke, P. (1986). Acoustic adaptations for parent–offspring recognition in swallows. *Experimental Biology*, 45, 179–193.

Bennett, J. F. (1976). *Linguistic Behavior*. London: Cambridge University Press.

Cheney, D. L. & Seyfarth, R. M. (1991). Truth and deception in animal communication. In C. E. Ristau, ed., *Cognitive Ethology: The Minds of Other Animals. Essays in Honor of Donald R. Griffin*. Hillsdale, NJ: Lawrence Erlbaum, pp. 127–151.

Chomsky, N. (1959). Review of B.F. Skinner, Verbal Behavior, New York, 1957. *Language*, 35, 26–58.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. New York: Praeger.

Darwin, C. (1871). *The Descent of Man, and Selection in Relation to Sex*. New York: D. Appleton.

Darwin, C. (1873). *The Expression of the Emotions in Man and Animals*. New York: D. Appleton (London: John Murray, 1872).

Dawkins, R. & Krebs, J. R. (1978). Animal signals: information or manipulation. In J. R. Krebs & N. B. Davies, eds., *Behavioral Ecology: An Evolutionary Approach*. Oxford: Blackwell, pp. 282–309.

Dennett, D. C. (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

De Veer, M. W. & Van Den Bos, R. (1999). A critical review of methodology and interpretation of mirror self-recognition research in nonhuman primates. *Animal Behaviour*, 58, 459–468.

Dingle, H. (1969). A statistical and information analysis of aggressive communication in the mantis shrimp *Gonodactylus bredini*

Manning. *Animal Behaviour*, 17, 561–575.

Endler, J. A. & Basolo, A. L. (1998). Sensory ecology, receiver biases and sexual selection. *Trends in Ecology & Evolution*, 13, 415–420.

Falls, J. B. (1982). Individual recognition by sounds in birds. In D. E. Kroodsma & E. H. Miller, *Acoustic Communication in Birds, Vol. 2*. New York: Academic Press, pp. 237–278.

Gallup, G. G. (1970). Chimpanzees: Self-recognition. *Science*, 167, 86–87.

Gallup Jr, G. G., Anderson, J. R. & Platek, S. M. (2011). Self-recognition. In S. Gallagher, ed., *Oxford Handbook of the Self*. Oxford: Oxford University Press, pp. 80–110.

Godard, R. & Wiley, R. H. (1995). Individual recognition of song repertoires in two wood warblers. *Behavioral Ecology and Sociobiology*, 37, 119–123.

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

Guilford, T. & Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal Behaviour*, 42, 1–14.

Haldane, J. B. S. & Spurway, H. (1954). A statistical analysis of communication in 'Apis mellifera' and a comparison with communication in other animals. *Insectes sociaux*, 1, 247–283.

Hazlett, B. A. & Bossert, W. H. (1965). A statistical analysis of the aggressive communications systems of some hermit crabs. *Animal Behaviour*, 13, 357–373.

Hogan, J. A., Kruijt, J .P. & Frijlink, J. H. (1975). 'Supernormality' in a learning situation. *Zeitschrift für Tierpsychologie*, 38, 212–218.

Hubel, D. H. & T. N. Wiesel (1963). Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *Journal of Neurophysiology*, 26, 994–1002.

Kant, I. (1961). *Prolegomena to Any Future Metaphysics* (transl. P. Carus), in particular Sections 23–24. Chicago: Open Court Publishing. [1st edn in German 1783, 1st edn thus 1902]

Krebs, J. R. & Dawkins, R. (1984). Animal signals: mind-reading and manipulation. In J. R. Krebs & N. B. Davies, eds., *Behavioral Ecology: An Evolutionary Approach*, 2nd edn. Oxford: Blackwell Science, pp. 380–402.

Lorenz, K. (1966). *Evolution and Modification of Behavior*. London:

Methuen.

Lynn, S. K., Cnaani, J. & Papaj, D. R. (2005). Peak shift discrimination learning as a mechanism of signal evolution. *Evolution*, 59, 1300–1305.

Marler, P. (1961). The logical analysis of animal communication. *Journal of Theoretical Biology*, 1, 295–317.

Marler, P. (1984). Song learning: innate species differences in the learning process. In P. Marler & H. S. Terrace, eds., *The Biology of Learning*. Berlin: Springer Verlag, pp. 289–309.

Marler, P. & Peters, S. (1988). The role of song phonology and syntax in vocal learning preferences in the song sparrow, Melospiza melodia. *Ethology*, 77, 125–149.

Maynard Smith, J. & Harper, D. G. C. (2003). *Animal Signals*. Oxford: Oxford University Press.

Medvin, M. B., Stoddard, P. K. & Beecher, M. D. (1993). Signals for parent–offspring recognition: a comparative information analysis of the calls of cliff swallows and barn swallows. *Animal Behaviour*, 45, 841–850.

Mitchell, R. W. & Anderson, J. R. (1997). Pointing, withholding information, and deception in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, 111, 351–361.

Møller, A. P. (1988). False alarm calls as a means of resource usurpation in the great tit Parus major. *Ethology*, 79, 25–30.

Morris, C. W. (1955). *Signs, Language, and Behavior*. New York: G. Braziller.

Ogden, C. K. & Richards, I. A. (1923). *The Meaning of Meaning : a Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: K. Paul, Trench,

Trubner. Owings, D. H. & Morton, E. S. (1998). *Animal Vocal Communication: A New Approach*. Cambridge: Cambridge University Press.

Pfeifer, J. (2006). The use of information theory in biology: Lessons from social insects. *Biological Theory*, 1, 317–330.

Rendall, D., Owren, M. J. & Ryan, M. J. (2009). What do animal signals mean? *Animal Behaviour*, 78, 233–240.

Ryan, M. J. (1998). Sexual selection, receiver biases, and the evolution of sex differences. *Science*, 281, 1999–2003.

Saussure, F. de (1959). *Course in General Linguistics*, C. Bally, A.

Sechehave & A. Riedlinger, eds. [transl. W. Baskin]. New York: Philosophical Library [1st edn in French 1916].

Searcy, W. A. & Nowicki, S. (2005). *The Evolution of Animal Communication: Reliability and Deception in Signaling Systems.* Princeton: Princeton University Press*.*

Seyfarth, R. M., Cheney, D. L., Bergman, T. et al. (2010). The central importance of information in studies of animal communication. *Animal Behaviour*, 80, 3–8.

Shannon, C. E. (1948). The mathematical theory of communication, I and II. *Bell System Technical Journal*, 27, 379–423, 623–656.

Shannon, C. E. & Weaver, W. (1963). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Sherrington, C. S. (1906). *The Integrative Action of the Nervous System*. New York: Scribner.

Smith, W. J. (1980). *The Behavior of Communicating: An Ethological Approach*. Cambridge, MA: Harvard University Press.

Steinberg, J. B. & Conant, R. C. (1974). An informational analysis of the inter-male behavior of the grasshopper *Chortophaga viridifasciata*. *Animal Behaviour*, 22, 617–627.

Tanner, D. A. & Visscher, P. K. (2010). Adaptation or constraint? Reference-dependent scatter in honey bee dances. *Behavioral Ecology and Sociobiology*, 64, 1081–1086.

ten Cate, C. & Rowe, C. (2007). Biases in signal evolution: learning makes a difference. *Trends in Ecology & Evolution*, 22, 380–387.

Tinbergen, N. (1951). *The Study of Instinct*. Oxford: Clarendon Press.

Towne, W. F. & Gould, J. L. (1988). The spatial precision of the honey bees' dance communication. *Journal of Insect Behavior*, 1, 129–155.

Weidenmüller, A. & Seeley, T. D. (1999). Imprecision in waggle dances of the honeybee (*Apis mellifera*) for nearby food sources: error or adaptation? *Behavioral Ecology and Sociobiology*, 46, 190–199.

Whiten, A. & Byrne, R. W. (1988). Tactical deception in primates. *Behavioral and Brain Sciences*, 11, 233–273.

Wiley, R. H. (1973). The strut display of male sage grouse: a 'fixed' action pattern. *Behaviour*, 47, 129–152.

Wiley, R. H. (1994). Errors, exaggeration, and deception in animal communication. In L. Real, ed., *Behavioral Mechanisms in Evolutionary Ecology*. Chicago, IL: University of Chicago Press, pp. 157–189.

Wiley, R. H. (2006). Signal detection and animal communication. *Advances in the Study of Behavior*, 36, 217–247.

Wiley, R. H. & Wiley, M. S. (1977). Recognition of neighbors' duets by stripe-backed wrens *Campylorhynchus nuchalis*. *Behaviour*, 62, 10–34.

Wilson, E. O. (1962). Chemical communication among workers of the fire ant Solenopsis saevissima (Fr. Smith). 2. An information analysis of the odour trail. *Animal Behaviour*, 10, 148–158.

Wittgenstein, L. (1968). *Philosophical Investigations*, 3rd edn (translated by G. E. M. Anscombe). New York: Macmillan. [1st edn 1953]

Zahavi, A. (1977). The cost of honesty. *Journal of Theoretical Biology*, 67, 603–605.

Zahavi, A. & Zahavi, A. (1997). *The Handicap Principle*. Oxford: Oxford University Press.

Summary and Final Thoughts

These chapters have explored, in a new way, how organisms evolve to perceive, communicate, socialize, and develop. Perception and communication underlie most behavior, of animals, plants, or microbes. As a result of noise, perception and communication become inherently unpredictable. Not completely unpredictable, but inevitably unpredictable to some degree. Therefore, so does behavior. Responses by every individual are unavoidably affected by noise. This final chapter reviews some of these fundamental conclusions and considers their deep significance for humanity. Before continuing, an important extension of these principles needs emphasis.

Noise in the Development of Individual Organisms

Chapter 6, Evolution by Natural Selection, makes the point that the development of an organism throughout its life depends on an interaction between its current state, ultimately derived from its genome, and its immediate environment. The development of each individual thus requires continual responses to environmental stimulation. It is an extended form of perception. Like all perception, developmental mechanisms evolve in inescapable noise.

Noise requires a new approach to developing organisms. A *tabula rasa* for development cannot work in the presence of irrelevant stimulation. To develop with as much predictability as possible, an organism must start with some inborn plans, insured by the stability of its genes and “canalized” for its early environment. An influx of noisy, extraneous stimulation cannot alone assure much success. Regularity in development requires some suggestions for where to start learning or responding. Each organism must include some predispositions to initiate appropriate responses to the environment. Evidence shows, for instance, that newly hatched songbirds have such predispositions for recognizing species-specific features of their songs. We can expect something similar for children's early preferences, including a preference for learning a human language.

Natural selection might thus explain why humans have neural mechanisms that provide a capacity for developing a language. It favors individuals that have this capacity. On the other hand, which language each individual learns depends ultimately on environmental stimulation. Natural selection might also explain a human predilection for imitating the behavior of parents, friends, or other associates. These neural mechanisms might predispose humans to associate with a few groups of recognized individuals and to match (to a degree) their language and practices. Natural selection thus might directly produce a predisposition to culture or language and only indirectly the specific rituals or language that are acquired from social interactions.

Variation in the development of individuals within any population also depends on interactions between each individual's genome and continuing stimulation from the environment. As a result, individuals might develop different personalities or different anatomies in response to environmental stimulation. Certainly training of the appropriate sort can, to some degree, alter artistic, scholastic, or athletic capabilities. Therapists or gurus might affect personality. These developmental influences from the environment are responses to stimulation. Like all forms of perception, they are inescapably affected by noise and thus are unpredictable to some degree.

Noise in Human Communication

When thinking about human communication in noise, it is important to keep in mind what is meant by the terms "noise" and "receiver's errors". A common feature of all scientific discourse is to convert ordinary words into specific technical terms. The words "cell" and "energy" are just two of numerous examples. In the context of Information Theory and the chapters of this book, "noise" means errors by receivers when responding to signals. The criterion for noise is errors by receivers. Noise is measured by counting such errors. Errors result because a receiver only knows whether an incoming stimulus reaches the receiver's criterion for a response or not. Such a stimulus might result in a correct detection or a correct rejection by the receiver. Yet it might also result in a false alarm or a missed detection. Whenever a receiver is aware of its sensory input, it only knows two possible states of the world, but actually there are four. Two result in correct responses, two result in errors.

An "error" in the context of Natural Selection is a response that has long-term disadvantages for the organism, in particular a response that results in leaving fewer copies of the organism's genes to future generations. It is a response that decreases the individual organism's eventual survival, its eventual reproduction, or its eventual effects of the reproduction of close relatives likely to share copies of the individual's genes. For humans, these effects of behavior are very difficult to measure. They are perhaps best related to the long-term well-being of an individual and its relatives.

It is also important to realize that "noise" as just defined by a receiver's errors does not imply that it is the receiver's fault. A receiver's responses can be errors for many reasons. They can result from problems of the signaler as well as problems of the receiver, both of whom have imperfect nervous systems that result in mistakes during the production as well as the reception of signals. In everyday language, we recognize that misunderstandings can arise because of mistaken saying as well as mistaken hearing. Problems can also arise in the environment during the transmission of signals. Masking of signals by similar energy in the environment is the way we usually think of noise in everyday life. Masking can result from the general level of energy in the environment and also from nearby individuals conducting communication irrelevant to the receiver. For spoken language, it might be the sounds of a nearby machine or a nearby conversation by other people. Another form of environmental noise is deception, misleading signals produced by individuals other than an expected signaler.

Any of these forms of noise produces errors by any receiver, including any human. The principles derived in these chapters apply to all forms of communication. The same principles apply to human languages.

In all of these processes, organisms respond to stimulation in the form of signals. Natural selection acts in every case to reduce errors in responses to stimulation. Errors as defined in these chapters are responses that are disadvantageous for the survival and reproduction, for the well-being, of each organism. Although natural selection tends to reduce errors, these chapters show that complete elimination of errors does not occur. Development, perception, and communication are inherently noisy. Perfection is not expected.

The Biology of Philosophy

A claim that perception is noisy opens some old philosophical questions. Our minds, our bodies, our environments have new relationships. Human minds can understand the mathematics of natural selection, which in its basic concept is just arithmetic. Natural selection, as thus understood, provides clear explanations for many deep aspects of human minds, even self-consciousness, free decision, and ethical behavior. Minds thus understand how natural selection produces minds. We can understand the biology of philosophy, as well as the more conventional philosophy of biology.

Ancient philosophers around the world, indeed everybody who stops to reflect, have noticed that what we see is not necessarily what is in front of us. Perception is potentially inaccurate. Obviously, we cannot count on perception alone to lead us to the truth. From this realization, some thinkers have concluded that our minds can know truth, but not by means of perception. Plato's heritage has a very long shadow. Descartes too recognized that misleading perceptions could not guarantee truth about the world, but he nevertheless argued that the world could be known. After all, he invented a crucial way to apply mathematics to the world, by combining algebra and geometry. In the end, he concluded, God must exist. That was enough to explain the mathematical regularity of the world, despite our imperfect perceptions. Kant likewise recognized the unreliability of perception despite the incontrovertible truth of mathematics. He supposed that our minds comprehended inerrant mathematics as a special category (or several categories) of thought. These special categories for mathematics were then extrapolated to categories for ethical behavior and for apprehension of beauty. Descartes and Plato had also used arguments for finding truth to explain ethics.

This quick summary of the philosophical issues of perception and stimulation is no doubt inexcusable. Kant for instance took four (at least) large books, all of them untranslatable and incomprehensible in detail, to explain his basic conclusions. Two points are nevertheless conspicuous in my lightning review. There have been two issues never considered – noise and natural selection. Instead of noise, errors are foisted either on minds or on the environment alone. Instead of natural selection, supernatural forces are conjured to produce the mathematical patterns of the world and of living organisms with minds.

Within the past century, three processes have been described by thoroughly vetted mathematics -- natural selection, noise in communication, and noise in perception. In each case, the basic math is not complicated. Some elaborations of the theories can look daunting. Yet each derives from arithmetic.

The approach in this book accepts this approach. If we accept arithmetic, we can understand natural selection of minds. Natural selection of communication is a process of co-evolution, the simultaneous natural selection of two kinds of behavior – signaling and receiving. Every communication involves two individuals, a signaler and a receiver. Each might send signals and respond to signals in different ways as situations change. Populations though should evolve to include individuals that send and receive signals as effectively as possible.

The chapters in the first section explain how this co-evolution works. Sending and receiving signals in noise evolves to a joint optimum. The sender uses just enough exaggeration in its signals, and the receiver sets its threshold just high enough. The result is an optimum. It is not perfection. The presence of noise creates trade-offs for signalers and receivers. The approach to perfection produces diminishing returns and augmenting costs. At the optimum, each party does the best it can, provided the other does likewise.

Communication can never be expected to be perfect. Chapter 8, Signal Detection and Communication, reviews the sorts of tactics that can improve the reliability of communication. Yet we cannot expect them ever to result in perfection.

Along the way, other conclusions have come to the surface. Noise in communication and in perception requires that organisms must make decisions whenever they attend to external stimulation. The mechanisms vary. Plants, unicellular organisms, and sea jellies make decisions without brains, even simple ones. Noise requires decisions, because a receiver or perceiver only knows two states of the world at a time -- either its sensory mechanisms have reached a criterion for response or not. Yet there are actually four possible outcomes each time. Two outcomes are correct – the receiver responds when an appropriate stimulus is actually present (mixed with noise) or it does not respond when the stimulus is absent and only noise is present. Two outcomes are errors – the receiver responds when only noise is present, or it fails to respond when a signal is present (with noise). The mathematics for optimal decisions is now also well established – Decision Theory assigns a utility to each alternative to calculate the maximal utility for each individual.

The inevitable presence of noise has other deep implications as well. For instance, noisy perception explains the ancient conundrum of imperfect perception. Combined with mechanisms for associative learning and categorization, it is easy to see how any such organism could acquire self-consciousness. Also the hypothesis of free will appears in a new light. Decisions are unavoidable for any organism in the process of noisy perception or communication. Each of us can learn from experience that our decisions include errors. We can adjust our thresholds for responses in order to minimize errors, but the contingencies of the universe are so vast that we cannot expect error-free decisions. Even in a completely deterministic macroscopic universe, our decisions are under-determined in the presence of incalculable noise. Our brains are responsible for our decisions. Decisions are also necessary about other people's responses.

Underlying all of this discussion is the operation of natural selection. The co-evolution of exaggeration in signals and thresholds for response, the evolution of mechanisms for associative learning and categorization of stimulation, the predispositions for learning highly specific tasks, all are subject to natural selection. No supranatural explanation is needed for any of these mechanisms. Mathematical natural selection does it all.

Even ethics results from natural selection, as indicated in Chapters 1, Evolution of Communication, and 6, Evolution by Natural Selection, where the evolution of cooperation is considered. A golden rule of ethics, "Do unto others as you would have them do unto you", can result from natural selection in the right circumstances. Natural selection explains the situations that promote such reciprocity (even, in special situations, altruism) and also the limitations of reciprocity. Again noise complicates evolution by natural selection. Cooperation, after all, relies fundamentally on communication. Noise results in imperfect communication and thus would sow imperfection in cooperation as well.

Because natural selection has such a pervasive influence on the evolution of organisms, including on communication and thought of all sorts, natural selection needs close examination. Chapter 6 reviews arguments for the action of natural selection. In the end there is a strange circuit in the argument. Human minds can understand natural selection. This book makes the case by reducing the actions of natural selection on communication and perception to mathematics. The result is a realization that natural selection can explain the evolution of minds with self-awareness, free decisions, and thought complex enough to understand natural selection itself. Natural selection thus produces a biology of philosophy.

This argument is not circular in a logical sense. It does not propose that natural selection produces minds that produce natural selection. It instead shows that natural selection produces minds that can understand natural selection. No matter what minds might be capable of doing, they are always subject to natural selection. It is plain arithmetic.

Hope and Humility

Realizing the pervasive influence of noise in all communication, can we expect humans to achieve perfection? Noise in human communication can take many forms, including all sorts of mistakes, misunderstandings, misinformation, and deception. Any and all noise, as these essays have explained, is to some degree unavoidable. Is there then hope for universal love or peace on earth? When we reflect that we are all signalers and receivers – our family, our best friends, our worst enemies, every writer and reader, every law-breaker, every teacher, every employer and employee, every leader, every politician, every soldier and general, indeed ourselves – the inevitability of noise in communication does not make universal love or peace likely in the long run. If perfection in communication is impossible, does the argument of this book have an optimistic ending? We should expect some success in communication, but should also be prepared for failures.

What path then might promote peace on earth? When confronting this question of an uncertain ending, Paul recommended, "Faith, hope, and love." Love, in Paul's sense of brotherly love, is crucial. It is after all the goal. Hope is also important. We should, all of us, never cease to try to understand each other, despite incessant error. We should attend to the conditions that minimize errors, by enhancing the contrast, redundancy, and familiarity of signals.

Hope and love, however, as vital as they are, do not seem to be enough. For Paul, as it has been for many others, the ultimate solution was faith. Augustine though was closer to the mark. He recognized what was still missing when he suggested, "Humility turns humans into angels". Only humility can usher us past error. The best hope lies in Gandhi’s confession, slightly revised, "Like any other fellow mortal, I confess to be liable to err. I hope, however, that I have humility enough to correct my errors". That still won’t get us to perfection, but it might be close enough. In line with Gandhi’s hope, I say, “Me too.” Shouldn’t we all aspire to humility enough to transcend noise?

Acknowledgments

My approach to investigating the evolution of communication has developed over the past several decades as a result of discussions with many students and colleagues in the Department of Biology at the University of North Carolina, in the Triangle Behavior Seminar (UNC, Duke University and NC State University), and in a course jointly taught in noisy conditions for a decade with Steve Nowicki at Duke University. I especially thank many current and former colleagues for discussions on the ideas presented here, but especially Lori Wollerman, Jean Boal, Marc Naguib, Jordan Price, and David Luther. Carl Gerhardt, Laura Miller and Maria Servedio provided helpful suggestions for parts of the manuscript. As any reader must realize by now, and as I acknowledge, there must be errors, for which I accept entire responsibility. My wife Minna was the sole source of continuous support throughout this project and many other endeavors. With humility, I avow she is my love and peace.

The chapters of this book, reprinted here with permission, with corrections and minor modifications, were first published as follows:

Encyclopedia of Evolutionary Psychological Science, T. Shackelford & V. Weekes-Shackelford (Eds.). 2021. Cham, Switzerland: Springer International Publishing. Chapters 1-4 and 6.

How noise determines the evolution of communication, *Animal Behaviour*, Vol. 124, 2017, Pp. 307-313. Chapter 5.

Receiver-signaler equilibrium in the evolution of communication in noise. *Behaviour (Special Issue in Honor of R. H. Wiley*, M. Naguib & J. Price (Eds.), Vol. 150, 2013, Pp. 957-993. Chapter 7.

Signal detection and animal communication. *Advances in the Study of Behavior,* Vol. 36, 2006, Pp. 217-247. Chapter 8.

Animal communication theory: information and influence, U. Stegmann (Ed.), 2013. Cambridge Univ. Press. Pp. 113-129. Chapter 9.