

Dealing with deception in biology

Carl T. Bergstrom

Department of Biology

University of Washington

Seattle, WA, USA

and

Santa Fe Institute

Santa Fe, NM

March 21, 2008

Life gets interesting when it gets complicated — and nothing complicates life so much as society. Flowing through the network of social conventions, institutions, obligations, and expectations that make civilization possible are a complex suite of individual motivations, sometimes common and sometimes conflicting, upon which tenuous alliances are formed and broken. We humans are fascinated by this complexity and drawn naturally to it. Here gossip seeps through the bedrock of friendship, here the politics draw us in against our better judgement, here theatre becomes gripping as intrigues multiply, here Walter Scott's tangled web demands every iota of our cognitive

ability to carry out our deceptions while unveiling those of others.¹ Society becomes possible through trust, coordination, and communication; this matrix of trust, coordination, and communication provides countless alcoves in which deception lurks and threatens to undermine the very social order that brings it into being.

Biology is no different. Biology gets interesting when simple structures aggregate and diversify to form larger and more complex units of organization. Cells bring together once-independent organelles, bodies aggregate billions of cells, colonies collect together thousands of bodies, and ecosystems comprise a webwork of competing, cooperating, and co-existing species. The diversity and complexity of the biological world emerges from these hierarchies of social organization — and from the intricate mechanisms required to keep such from splintering along lines of individual incentive.

Such is the central thesis of the influential 1995 monograph, *The Major Transitions in Evolution* [1], which laid out a research program for a generation of evolutionary biologists. Its authors, John Maynard Smith and Eors Szathmary, envision the history of life as the story of a series of major evolutionary innovations and transitions allowing organisms and societies to increase in complexity and efficiency [2]. Some of these transitions, such as the shift from unicellularity to multicellularity, or the shift from asociality to eusociality, facilitated cooperation by aggregating previously independent agents into ensembles with linked reproductive fates. These transitions allowed the agents to exploit economies of scale, and in particular, gains to specialization. Other transitions engendered cooperation and trust by changing the rules of the games without physically aggregating the individuals in-

volved. In these latter transitions the agents impose upon one another the strategic incentives which allow them to continue to operate independently, yet once again facilitate the economies of scale and gains to specialization that come with cooperation and coordination.

What does this have to do with deception? A great deal. First, to benefit from aggregation or cooperation of the sort discussed above, participants need ways of coordinating their actions. And for coordination, they need communication. This opens the door for deception; where one can communicate, one can manipulate [3]. Somehow the deception problem must be overcome to move through the major transitions that Maynard Smith and Szathmary contemplate.

Second, when it comes to cooperation, one of the first things to share is information. A key insight that Maynard Smith and Szathmary provide is the idea that most if not all of these transitions increase the scale upon which organisms and societies can acquire, store, process, and transmit *information*. Innovations, such as the shift from RNA-based to DNA-based genetic information, or the shift from simple signals to combinatorial representational language, offer leaps in information technology that allow organisms to better extract information from their environments and transmit this information to other individuals. Why is information and information-sharing so important in generating major evolutionary transitions? Cooperation and sharing are particularly likely to emerge for information resources rather than physical goods, because of the unique stoichiometry of information. As Lachmann and colleagues [4] showed in an elegant model of information acquisition and exchange, information sharing works differently than the sharing of physi-

cal resources. A quotation commonly attributed to George Bernard Shaw summarizes the underlying concept that drives their model:

If you have an apple and I have an apple and we exchange apples then you and I will still each have one apple. But if you have an idea and I have an idea and we exchange these ideas, each of us will have two ideas.”

—George Bernard Shaw

Those readers whose political inclinations leave them skeptical of sermons on sharing delivered by celebrated socialists may consider Thomas Jefferson to be a more reputable source:

He who receives an idea from me, receives instruction himself without lessening mine; as he who lights his taper at mine, receives light without darkening me. That ideas should freely spread from one to another over the globe, for the moral and mutual instruction of man, and improvement of his condition, seems to have been peculiarly and benevolently designed by nature, when she made them, like fire , expansible over all space, without lessening their density in any point.

—Thomas Jefferson, 1813²

Whatever one’s leanings, there is clearly a highly favorable stoichiometry for sharing information. But where receiving shared information can be beneficial, it can be a dangerous proposition as well. Signalers have the means

to effect changes in the behavior of those to whom they signal. Where the interests of signaller and signal receiver diverge, there exist both incentives and opportunity for manipulation by sending misleading information. Deception is the major obstacle to information sharing. And the living world is rife with deception. From the lure that an anglerfish uses to attract prey [5] to the false alarm that a flycatcher raises to dissuade competitors [6], from bluegill sunfish males that sneak matings by masquerading as females [7] to the mimic octopus that can imitate a wide range of poisonous creatures and other underwater objects [8], from the false mating signals of carnivorous fireflies [9] to the sham regenerated claw of a fiddler crab [10], from the chemical mimicry that caterpillars use to invade the nest chambers of ants [11] to the bluffing threats of a molting stomatopod [12], organisms deceive one another in every imaginable way in order to attain every conceivable advantage.

So this creates a puzzle. On one hand, communication provides organisms with the means to deceive, and indeed deception is common in animal signals. On the other hand, for animal communication systems to evolve, they have to be beneficial — and thus presumably somewhat honest — on average. After all, if they were not beneficial, the intended signal receivers would evolve to ignore them. And if signal receivers ignored these messages, they would be useless, and signallers would eventually evolve not to send them. Maynard Smith and Harper put this in game-theoretic terms: “it is not evolutionarily stable for the receiver to alter its behaviour [in response to a signal] unless, on average, the signal carries information of value to it.” [13]. This insight can be formalized within game theory using the concept of information value [14].

To resolve this puzzle, let us explore the ways in which living organisms deal with all of this deception. In order to do so clearly, we will distinguish between two different forms of deception that organisms must deter or detect.

1. **Deception by society members.** The “legitimate participants” in a social interaction or signaling situation have different interests from one another, and thus have incentives to manipulate one another by deception³ .
2. **Subversion by rogue outsiders.** The legitimate participants in a social interaction or signalling system have coincident interests, but “rogue outsiders” may attempt to parasitize the systems by subterfuge.

We might see the former type of deception in animal (or human!) courtship. For example, when a suitor displays to a potential mate, both are legitimate participants in the interaction, but the suitor has an incentive to over-represent his or her quality in an effort to impress. We might see the latter type of deception when a caterpillar mimics the hydrocarbon signals that ants use for nest-mate recognition, in order to gain access to the brood chambers and a ready meal of ant larvae. We will explore further examples of both types in the subsequent sections.

Before we do that, we can draw out the comparison by taking a slight detour to consider a recent human institution. The internet auction site eBay has been remarkable successful at extending the scope of small-scale interpersonal commerce from local to global. As eBay draws upon explicitly in their business model, this requires the creation of trust within the community, and requires that there be effective mechanisms in place to deal with the threat

of deception [15, 16]. Think about the various perils that you might confront when you shop for a rare book on eBay. An otherwise honest dealer might palter, accurately listing a book as “First edition,” but omitting the addition detail that this particular volume was printed in such great quantities as to render the first edition nearly worthless. In a more serious deception, an unscrupulous antique dealer might exaggerate the condition or quality of a rare book in order to generate a higher selling price.

These are examples of deception by society members. Both the dealer and the potential buyer are the intended participants in the auction system; the only problem is that the dealer has an incentive to deceive so as to increase his profits. The eBay system deals with this threat in a number of ways: with an extensive reputation system, with a suite of dispute resolution tools and procedures, and with limited third-party guarantees over some transactions.

Compare this to deception by rogue outsiders, such as the “phishing” schemes that Paul Thompson describes in detail later in this volume. For example, a perpetrator might set up a fake website designed to look like eBay, and send out emails to eBay users in an attempt to lure them to the site where they may be tricked into entering credit card information or other valuable data. In this case, the perpetrator is not an intended participant in the interaction; eBay was not put into place so that con artists could phish for credit card numbers. Thus the problem in this situation is that the con uses deception to insert herself into the communication flow between the intended parties. And eBay goes to extensive lengths to help users protect themselves against this form of deception as well. To help users avoid phishing scams, eBay provides a tutorial on these deceptive tactics and how to avoid falling

prey to them [17]. Moreover, eBay communicates with users only through a highly protected channel, its own site-internal message system, and never by general email which can more easily be spoofed.

So that is eBay; numerous other internet communities and, more generally, human social institutions face similar problems. But what about biology? Where in non-human biological systems do these opportunities for deceit arise, and what sorts of mechanisms have evolved to deter deception?

Deception by society members

The basic problem of honest communication is as follows.

Information asymmetry: A *signaler* has private information that a *signal receiver* does not.

Gains to exchange: They could both gain if they could honestly share this information.

Strategic conflict: The signaler and receiver have different preferences with regard to the receiver's actions — and thus the signaler may have an incentive manipulate the receiver by means of deception.

How can honest communication be ensured?

From the profusion of signals that saturates the living world, we can infer that this problem is somehow resolved. Whether you are walking along rocky

ocean shore, bicycling along a forest path, or simply soaking up the sunshine in a mountain meadow, a huge fraction of your sensory experience comes from stimuli that evolved precisely for the purpose of operating as signals. Calls, patterns, colors, fragrances — these are just a few of the modalities by which signals are sent and received.

In the early 1970's, economist Michael Spence [18, 19] and biologist Amotz Zahavi [20, 21] independently proposed what was essentially the same solution to this problem. Spence proposed his solution to explain how higher education serves as a signal from employee to employer, and developed a formal game-theoretical model in its support; for this work he shared a Nobel prize with George Akerloff and Joseph Stiglitz. Zahavi proposed his solution in an effort to understand why animals often produce extravagant ornaments and displays, and his strictly verbal formulation was met skeptically [22, 23, 24] until formalized mathematically many years later by other researchers [25, 26, 27]. Before proceeding to a concrete example, let us summarize the general solution.

Spence-Zahavi solution: If signals are costly and lying costs more than honesty, everyone may do best by simply telling the truth.

But why would this be? Why would lies be more costly than honest signals? To answer this question, we turn to Zahavi's paradigm case, the tail of the peacock (*Pavo cristatus*)⁴.

In this example, the peacock is the signaler, and his perspective mate, the peahen, is the signal receiver. The cock has private information about

his own condition: he alone knows whether he is strong or weak, well-fed or undernourished, healthy or parasite-ridden. The hen would benefit from knowing this information, because she could then make a good choice as to whether or not to accept the peacock as a mate. (We assume that she benefits from choosing a high-quality male to father her offspring.) But the cock may have an incentive to deceive her. If he is in poor condition, he would do well to feign otherwise so as to avoid being rejected.

How can honest communication occur in this circumstance? The peacock's extravagant tail is the key. A weak or sickly male can scarcely afford to divert energetic resources from basic upkeep to the production of ornaments, and moreover he would have a hard time escaping from a predator if his flight were hindered by a long tail. A strong and healthy male, by contrast, can readily afford the additional costs of producing bright colors and a long tail, and moreover can usually escape a predator even when his flight is somewhat restricted by the length of his tail.

Because only the high-quality males can afford bright colors and long tails, peahens prefer mates with these characteristics. High quality males, for their part, produce these bright colors and extravagant plumes to ensure that they are chosen as mates by females. Low quality males cannot afford to do so, and so they will produce duller colors and shorter tails. Perhaps next year they will be stronger and able to be more ambitious with their plumage.

This was a wonderful idea on Zahavi's part, because it resolved two huge puzzles in evolutionary biology, by showing that each puzzle was actually the solution to the other. The puzzles are these: (1) Why are signals honest

despite incentives to deceive? And (2) why are so many biological signals extravagant if natural selection favors efficient use of resources?

But does Zahavi’s solution work? Game theoretic models indicate that it does. To show why with full rigor, we need to deal simultaneously with the signaling strategy of the signaler and the response strategy of the signal receiver and show that each is a best response to the other; Bergstrom et al. [28] provide a general methodology for doing so. Without going into the mathematical complexities of that approach, we can still capture about much of the intuition for why this mechanism works with a simple graphical model, after refs. [25, 29]. Here we treat the receiver’s response to different signals as fixed, and look at the properties of the optimal signaling strategy for signalers, given these responses.

Figure 1 shows a hypothetic set of fitness costs and benefits for peacocks playing the “mating game” described above. The dark concave-down curve labeled *fitness benefits* indicates the advantage — in terms of mating success — that accrues to a male as a function of the size of his tail. The lighter concave-up curves illustrate the fitness costs — in terms of energy expenditure, increased predation risk, etc. — of producing a tail, for peacocks of low, medium, and high quality respectively. High quality peacocks can produce larger tails more cheaply than can medium quality ones, who in turn can do so more cheaply than low quality ones.

Each peacock does best to maximize his fitness, and thus to choose a tail size that maximizes the difference between the fitness benefit received and the fitness cost of producing it. The dashed lines indicate the optimal tail sizes for the low, medium, and high quality peacocks; the dots indicate the

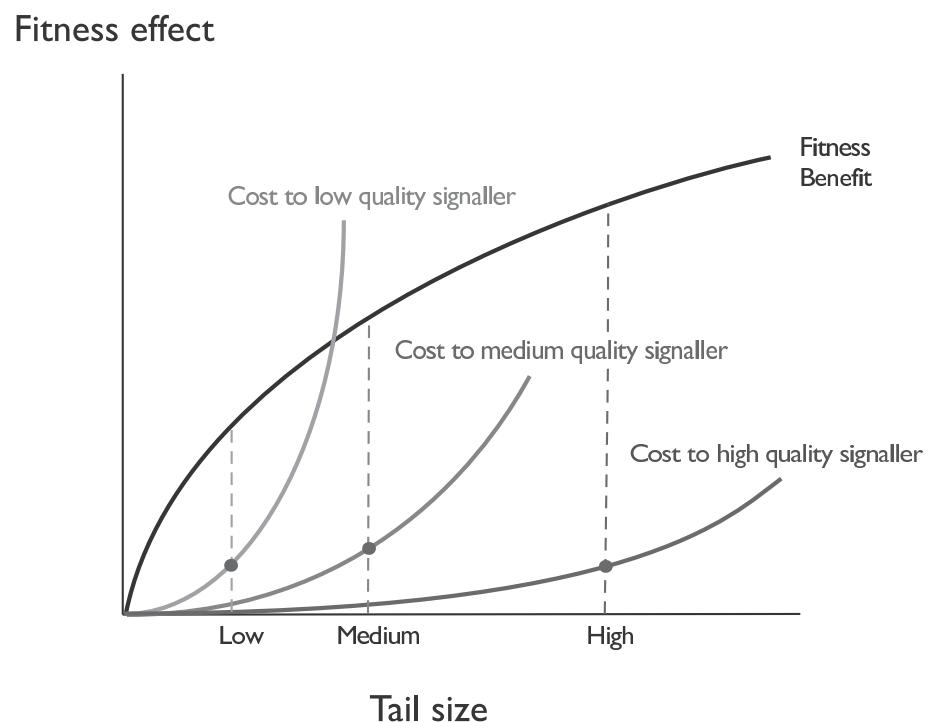


Figure 1: Costly signaling, after Lachmann et al. 2001

fitness costs incurred by these birds.

From this figure, we can see costly signaling in action. Even though peacocks are free to “choose” the optimal tail length given their condition, we see that here:

1. **Signals are costly.** Each cock incurs a non-zero fitness cost from producing its showy tail, and
2. **Signals are honest.** Higher quality birds produce larger tails, and the hen can thus infer quality from the size of the peacock’s tail.

We can view this diagram as an illustration of the mathematical vindication of Zahavi’s idea. Costly signals can provide a way of dealing with deception in biological systems.

But this seems to be an extremely wasteful way of transmitting information. Indeed, in some cases, costly signaling systems such as that above can leave signaler and signal receiver worse off than if no signal were sent at all [30]. Can there not be some way of sending honest signals at reduced cost?

To answer this question, we can take a cue from the house sparrow (*Passer domesticus*). This species (and many related species, including the often-studied Harris sparrow *Zonotrichia querula*) signal dominance or fighting ability using relatively minor and inconspicuous variations in plumage, such as variable throat bibs or forehead patches [31, 32, 33]. In the case of the house sparrow, the signal is the size of the black throat-patch. The size of the throat patch honestly signals fighting ability: birds with larger badges are less likely to be challenged and more likely to win if challenged than birds with smaller badges.

Fitness effect

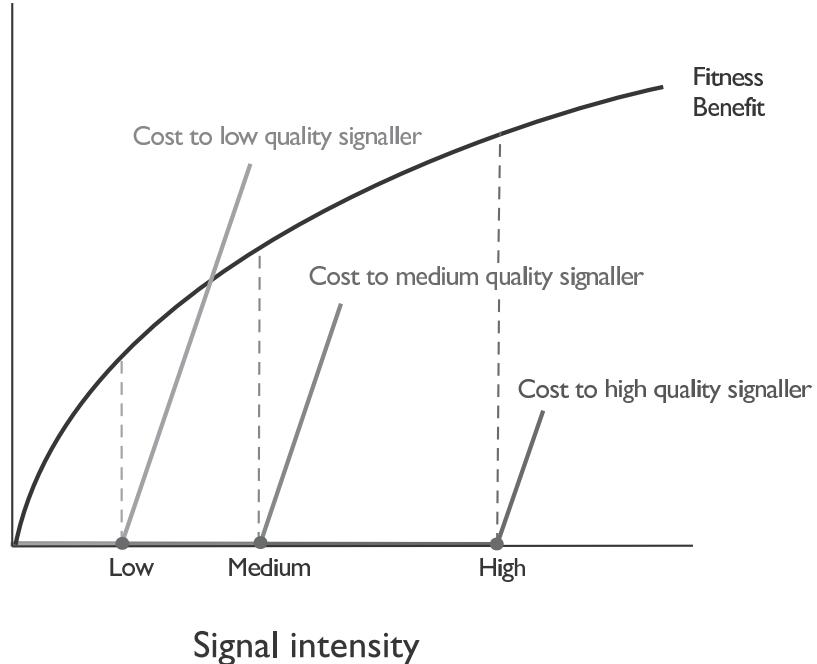


Figure 2: Cost-free signaling, after Lachmann et al 2001.

In the sparrow's case, in contrast to that of the peacock, the cost of actually *producing* the signal is very low: it is only the negligible expense of altering the shade of a few feathers. So what keeps this signaling system honest? Why do the lower-status birds not produce a deceptively large throat patch so as to feign dominance? The answer comes from the behavior of the other individuals in the social environment. Sparrows that "cheat" by exaggerating their own condition in their choice of badge are attacked and punished by conspecifics [31, 34, 35].

Here we see a different kind of honest signaling. Producing the signal is not costly in and of itself, but rather the costs only accrue through the actions

of other birds that enforce the signaling conventions, attacking any bird that produces too large a throat-patch for its status. When signal costs arise from the punishing behavior of the signal receiver, signals can be entirely free so long as they honest, and expensive only if deceptive [29]. Figure 2 provides a geometric representation of this type of honest signaling. Here signals are honest just as in the peacock example of Figure 1, but now the optimal choices of signal for each signaler, indicated by the shaded dots, are not costly. This figure illustrates that what enforces signal honesty is not the cost of signaling a particular quality, but rather the marginal cost of signaling that quality. In other words, it is not the expense of honest signals that enforces honesty, but rather the increase in expense that comes from sending dishonest signals.

Why does the peacock get stuck having to construct and maintain an elaborate and expensive tail in order to signal his quality, whereas the sparrow can signal quality with a simple and inexpensive adjustment to throat coloration? The difference is that in the peacock's case, there is no ready way for the signal receiver to verify the accuracy of the message. To figure out whether or not the cock really had the good genes that he was advertising, a peahen has to mate with him, raise her offspring, count the survivors among them, and do the necessary statistics to figure out whether her mate has deceived her as to the quality of his genome. By then, of course, he is long gone. In this system, there is no ready way for receivers to detect and punish misleading signalers, and so the cost necessary for honesty must come from signal production. The sparrow, on the other hand, can easily assess the honesty of a throat badge signal. All it has to do is to provoke a fight

with the signaler and see whether the signaler is indeed as tough as it has indicated by its coloration. Here, since assessment is quick and easy, signal costs can come from receiver behavior and need not be associated with signal production [29].

As we move beyond the types of animal signals described above to more complicated forms of communication such as combinatorial syntax and referential meaning that we see in human language, it becomes on one hand impossible to stabilize honest signals by assigning appropriate production costs to specific signals [29], and on the other hand a whole new suite of extended possibilities for deceptive communication arise [14]. Costs associated with deception will almost always come from the responses of receivers. Nonetheless, much of the basic logic illustrated in Figures 1 and 2 may continue to apply. For *Homo economicus* as well as *Pavo cristatus*, the choice of whether to deceive or not comes down to cost-benefit analysis between the expected benefits from sending a particular signal and the costs associated with doing so. A signaler will advance his cause until the marginal cost of pushing further exceeds the marginal benefit of doing so.

For human communication, much of this may be mediated through systems of reputation. If I lie to you, you may retaliate directly — but more likely you will simply adjust your assessment of my character, trustworthiness, and desirability as a partner be it in commerce, scholarship, or love. While human communication may require something like reputations in order to enforce honesty, human language also has the property of facilitating social enforcement by reputations. With complex referential communication, I am able to share information about those who have wronged me, and

gain information about others reputations even in the absence of first-hand experience.

In this section, we have seen how biological systems facilitate honest communication despite incentives to deceive. In short, the game-theoretic answer to how one deals with deception is this: *the structure of the communication “game” must include the appropriate strategic incentives against dishonesty.* These incentives may take the form of signal costs, reputation effects, or any of the other mechanisms — such as reciprocity, partner choice, and sanctions — that bolster cooperation [36, 37]. The game theoretic study of signaling draws its continued research interest from exploring at the myriad forms that these incentives can take, and from studying the dynamical processes that give rise to these forms. How can the appropriate incentives and disincentives evolve by natural selection (in the case of animal signaling) or be constructed by those designing institutions (in the case of many human communication systems) to deal with deception?

And with talk of constructing the appropriate incentives, we come back around to eBay. What eBay has done, and what allowed eBay to extend the social and geographic range of trusted commerce, was to set up highly efficient ways of distributing reputation information: a centralized, searchable database for each participant listing the experiences of all others who have previously dealt with that individual [15]. Compared to previous approaches of inquiring about another’s character through one’s own social network, this is a much more powerful way to get information about potential trade partners and their likelihood of acting deceptively. And thus eBay was able to capture the long tail of the market [38] without requiring the long tail of the

peacock as collateral against deception.

Now let us return to biology and look at how biological systems avoid deception by rogue outsiders. Thus far we have concentrated on an example in which one bird signals to another bird. As we shift our focus to deception by rogue outsiders, we will shift our attention from communication between animals to communication among the cells within a single animal. We will use the cells of the vertebrate adaptive immune systems as our paradigm “society” and consider pathogens such as bacteria and viruses as the rogue outsiders. In particular, we will look at the way in which immune cells, in their communications with one another, avoid being deceived by pathogens that may attempt to interfere.

In our discussion of signaling among intended participants, we have been able to lay out a clean and hopefully clear picture of the strategic conflicts that arise and how they can be resolved. This is in large part because the theory treated above is grounded upon a well-developed foundation of simple game-theoretic models that highlight the important aspects of the problem while abstracting away unnecessary details. The study of subversion by rogue outsiders has not yet reached this point; researchers have yet to identify the clear, simple models that capture the essence of the problem but include nothing more. In part, this is because the problem of deception by rogue outsiders has received less attention in the theoretical behavioral ecology and game theory literatures to date. In part, it is due to the difficulty in circumscribing the range of possible deceptions that an outsider can employ. As a result, the following section is somewhat more speculative and open-ended than was the preceding one. This is exciting: the discussion in the

next section describes an area that is ripe for future theoretical development.

Subversion by rogue outsiders

Think about the staggering challenge that our immune system faces in doing its job. To function properly, our immune system needs to be able to acquire information about what is self and non-self. It must continually scan the cells of the body for any sign of non-self. It has to process this information to mount appropriate responses. It must coordinate those responses among the millions of cells involved. And ideally, it should store this information as immune memory for subsequent use.

In doing all of these things, our immune system has to be extremely sensitive to detect rapidly-reproducing pathogens early on in an infection. It has to identify the proverbial needle of a non-self protein (e.g., one produced by a virus) among a haystack of self proteins within any individual cell. It must be incredibly broad so as to recognize and respond to any of practically countless variety of pathogens that could arise over evolutionary time. And yet every individual immune response must very narrowly target the pathogen or pathogen-infected cells, without attacking other parts of the body. Finally, our immune systems must have a very low rates of type-1 error (false positives), because any such mistake can trigger a potentially disastrous auto-immune reaction.

The inevitable tradeoffs between these requirements pose a control-theorist's nightmare, but this is only the beginning. To make matters immeasurably worse, pathogens typically replicate within the body of the host, and thus

have ample opportunity to subvert immune function by sabotage or subterfuge [39, 40]. For example, pox viruses such as chickenpox and smallpox have evolved not so much to avoid detection by the immune system, but rather to confuse the immune system and render it impotent [41]. Among other stratagems, they target the chemokine signaling molecules that the immune system uses to regulate and coordinate its responses. Pox viruses interfere with the chemokine signalling system in almost every imaginable way. They sabotage some chemokine signals by producing enzymes that degrade the signal molecules. They spoof other chemokine signals by producing false signal proteins that stimulate chemokine receptors. And they tamper with gene expression on the part of the host organism, altering chemokine signal production in that way. Pox viruses also sabotage the receptors, both by attacking them directly and by producing decoy chemokine receptors that attract the host’s chemokine signal molecules and prevent those signals from reaching their true targets.⁵

Due to the threat of this kind of deception, immune systems have to be robust not only to noise, but also to targeted misinformation and other forms of deceptive signals or information attacks: they have to be *strategically robust*. Another computer analogy can help to elaborate upon this distinction between robustness and strategic robustness. When constructing a mission-critical computer system, the engineers need to build in sufficient robustness to noise and accidental component failure that the system will function despite occasional power spikes and electrical interference. They need to make sure that it will keep working across a range of temperatures, and even when a few resistors burn out or when one of the memory chips fails. Designing

a strategically robust system is an even greater challenge; here one has to construct a computer system that will continue to function even if an expert technician with access to the inner workings of the machine tries jam it with targeted electrical interference or sabotage it with malicious software or by removing key physical components.

How can immune systems meet this challenge? How can they carry out the coordination and communication necessary to respond specificity, breadth, precision, and accuracy, *while simultaneously remaining robust to deceptive signals from pathogens*? As immunologists come to an increasingly refined understanding of the intricate molecular biology of immune learning, recognition, signaling, regulation, and memory, they are starting to uncover the answers to this question. And by comparing the tricks and tactics that the vertebrate adaptive immune system uses to avoid internal deception with the tactics used by other immune systems, evolutionary biologists can uncover general themes in the way that biological systems evolve to deal with deception from within.⁶

This type of study is still in its early stages in biology; we do not yet have a detailed theory of systems evolve to avoid internal deception and subversion. But to illustrate the sorts of insights that biology does offer, in the paragraphs that follow we will briefly describe just a few of the common mechanisms of avoiding internal deception and subversion, with examples drawn from the vertebrate adaptive immune system. A more detailed discussion and examples from other immune systems is provided in Bergstrom and Antia's 2006 paper [42].

Redundancy. One of the most straightforward mechanisms for ensuring

that a system continues to function smoothly despite noise, component failure, or sabotage and deception is to use multiple, redundant pathways to carry out the desired functions. For example, if one needs to ensure an uninterrupted source of power, it is very useful to have one or more backup generators in place in case something happens to the primary power source. If a system has k redundant pathways and any single pathway fails with probability p , the chance of all pathways failing — and thus the system as a whole failing — is the much smaller value p^k . We see this strategy of redundant pathways in the vertebrate adaptive immune system, which deploys multiple branches, including non-specific and other forms of innate immunity, cell-mediated immunity (killer T-cells), and humoral immunity (the antibody response) in its efforts to guard against invading pathogens. If one branch is shut down by sabotage or deception, the other branches remain and may well be sufficient to eliminate the threat. Having multiple redundant defenses also makes it harder for a pathogen to evolve ways around an immune system. Even if the pathogen manages to deceive one branch of the immune system, the others can eliminate it — and thus a pathogen gets little if any fitness benefit from outfoxing a single one of the immune pathways.

Distributed control. In systems where internal deception is not a concern, one the most effective ways to achieve coordination among multiple dispersed components is to have a *central controller* that instructs the many components of the system via *broadcast signals*. In living systems, we see precisely this approach in endocrine (hormonal) regulation. For example, the pituitary gland is a central controller that regulates numerous metabolic, developmental, and reproductive processes by emitting a suite of hormone signals that

control the behavior of numerous organs and tissues. But systems that are regulated by a central controller can be highly vulnerable to deception. If some antagonistic agent (e.g. a pathogen) take over the central controller to alter the nature and timing of signals, or even spoofs the broadcast signals, it can deceive the dispersed components that receive these signals, and alter the behavior of the entire system for its own purposes.

By way of analogy, suppose I want to coordinate the actions of a set of colleagues so that everyone submits a paper on the same day. One way to do this is to give each a small radio, and to broadcast a message “submit now!” at the appropriate time. But this method is vulnerable to subversion: someone could steal my transmitter, or spoof my message, and thereby alter the behavior of all of my colleagues in any way that she desired. If subversion is a concern, a safer alternative would be give each of my colleagues a wristwatch with a calendar, and let them check their own watches and then act at some prearranged time, e.g. noon on January 1st. Once a plan of this sort is put into place, there is no central target for an antagonistic agent and no single broadcast signal that the agent can spoof to take over control of the individuals’ actions. To change the behavior of the group, the saboteur would have to access and alter each and every wristwatch.

The vertebrate immune system, which functions to eliminate pathogen threats and thus is a natural target for disruption and deception by pathogens, makes scant use of central control and broadcast signals. Instead we see an extensive reliance on *distributed processing*. Decisions and commands are “pushed to the periphery”: sensing and control occurs over small local scales via signaling among the individuals components — immune cells — circulat-

ing throughout the body. While the immune system has multiple mechanisms for turning off immune reactions that are directed toward the self or that are ineffective, these are highly local in operation. There is no command center that a pathogen can take over or spoof in order to cancel an immune reaction with one simple signal. We see a similar logic of control in the intracellular immune system of RNA interference [42].

Commitment instead of feedback control. One of the foundations of control system design is the use of feedback or “closed-loop” control. A feedback controller measures the output or progress of a system, compares this output with the desired trajectory, and adjusts its input accordingly [43]. Feedback controllers allow systems to function effectively across a range of conditions, stabilize otherwise unstable dynamical processes, and facilitate closer tracking of the desired trajectory. Relative to open-loop control, in which there is no mechanism for monitoring a system’s progress and responding accordingly, closed loop control is a highly efficient way to regulate and coordinate behavior. Thus it is little surprise that feedback control is widely used in biological systems from adjusting the circadian clock to the control of metabolism, from cell signaling to hormonal regulation, from bacterial navigation to DNA replication to limb development [44].

But feedback control is risky in the face of potential internal deception. Implicit in the idea of feedback control is that the system responds to certain kinds of stimuli — and where these stimuli can be faked, the system can be coopted. In other words, when one cannot trust the controller, or one cannot trust one’s measurements of a system’s progress, feedback control can be dangerous.

Thus perhaps it is unsurprising that the vertebrate adaptive immune system omits feedback control in some of its regulatory machineries. In one of the most striking examples, recent experiments and mathematical models [45, 46] indicate that the growing populations of CD8 T-cells that deal with viral infections do not monitor the current density of pathogens within the body, contrary to the common belief. Indeed, the population of CD8 cells targeting a specific pathogen often continues to grow long after that pathogen is cleared by the immune system or by artificial means. Rather than relying on feedback control, CD8 cell lines appear to commit early the course of an infection to a “programmed” period of expansion that cannot be halted even if the pathogen disappears entirely. This makes sense; early in the course of an infection, a pathogen is at low density and is less likely to be able to tamper with immune sensing and signaling. Thus the cells of the immune system commit to a plan of action at this point, and despite some inefficiency from foregoing feedback control, they do not adjust this plan later, when the pathogen could potentially be at higher density and better able to tamper with immune regulation.

Cross-validation. Another way to avoid being tricked by subversive signals from rogue outsiders is to make sure that no single signal is sufficient to initiate a potentially dangerous course of action. If I want to make sure that I am not fooled into doing something by a single charlatan, I may require not only a primary message from a primary signaler, but also a set of supporting messages from other individuals. While closely related to the concept of redundancy, in which multiple independent systems back one another up in case one is subverted, cross-validation features a single system that requires

multiple inputs from varied sources before taking action.

For example, if an immune response could be cancelled by a single chemokine signal, the system would be highly vulnerable to deception by false chemokines of the sort that the pox viruses commonly produce. But if it takes several different signals of several different chemical classes to down-regulate an immune response, deception is less likely. If these multiple requisite signals are interrelated in complex ways (e.g., if one serves as a checksum for the others), deception by spoofing is even harder. For these reasons, it seems likely that cross-validation will be important in the structure of immune signaling systems.

The field of immunology, with its endless roster of cell types and signals, receptors and modulators, with its hopelessly entangled pathways of communication and regulation, is among all of the biological sciences perhaps the most notoriously hard to learn. One might conjecture, not entirely in jest, that this is no accident. Perhaps it has evolved to be so. The same cascades of complexity that stymie legions of medical students and fascinate generations of researchers may have evolved precisely because what students and researchers can learn quickly, pathogens can learn as well through the action of natural selection on trillions of virus particles reproducing many times a day within each of billions of host individuals over the time span of millions of years.

In this section, we have considered how one particular biological system, the vertebrate immune system, carries out the extensive information-gathering and signaling that it requires despite the threat of deception by rapidly evolving pathogens. The structures of other immune systems offer

comparable lessons [42]. To prevent deception by rogue outsiders, it is not sufficient to evolve or impose strategic incentives on the intender signaler; the intended signaler is not the one who is potentially causing the trouble. Rather, there can be an unending line of additional individuals trying to find ways in to exploit the system and they must be deterred as well. Perhaps there are no hard and fast rules about how to do this, but in examining the structure and control logic of immune systems, we do see repeated use of a few key design principles, including redundancy, distributed control, commitment, and cross-validation.

Society requires coordination and communication, which in turn requires honest communication among the participants in a social group. Therefore, to facilitate any sort of social structure and interaction, there has to be some way to deal with the threat of deception. Natural selection has been grappling with this problem for several billion years, innovating, testing approach after approach, solution after solution in the crucible of biological competition. Perhaps as we seek ways to deal with deception in our own societies, in our institutions, in our own communication systems, we can learn from what natural selection has devised, much as we have benefited from biologically inspired design in so many other areas of engineering.

Notes

¹“Oh what a tangled web we weave, When first we practise to deceive!” Sir Walter Scott, *Marmion*, Canto vi. Stanza 17.

²Letter: Thomas Jefferson to Isaac McPherson. 13 August, 1813, Pp. 333–334 in *The Writings of Thomas Jefferson*, vol. 13. 1903: Thomas Jefferson Memorial Association of the United States.

³The term “legitimate participants” is a shorthand for describing those individuals whose historical participation in the interaction was causally responsible for the emergence of the signaling system in the first place.

⁴ Zahavi’s original formulation of the handicap principle differs from its modern interpretation; we will follow the modern view here. The difference is this: in his original papers on the handicap principle, Zahavi viewed handicaps as credible because they cause natural selection to “screen” signallers more intensely. Carrying a handicap - and surviving nonetheless - serves as a statistical signal of strength. When biologists think about costly signals today, they typically assume instead that signaller makes a strategic choice of how large of a handicap to produce, taking into account both the cost of doing so, and the benefit that will come from the response of the signal receiver. This choice need not be a conscious decision or calculation on the part of the signaller; it can be a decision rule that is encoded by the genes and tuned through the action of natural selection. Relative to weak individuals, strong individuals can bear greater handicaps at lower costs, and so they will choose to produce larger ornaments. As a result, handicap size serves as a reliable signal of strength - and thus the receivers’ preference for large handicaps is justified.

⁵Not only do viruses practice trickery by spoofing immune signals and fashioning decoy signal receptors, but their very ability to do so as also been acquired by a sort of trickery. In many cases, the genes that the virus uses to tamper with a host’s immune system have been stolen from the host species’ genome at some earlier point in the virus’s evolutionary history. By virtue of replicating within host cells using the host’s genetic machinery, viruses have “access” to the full genomes of the host and can incorporate modified versions of host genes into their own viral genomes.

⁶The familiar vertebrate adaptive immune system is only one of many immune systems

that has evolved in the biological world. For example, bacteria rely on simple immune-like pathways known as restriction-modification systems to detect and destroy viral DNA. Many eukaryotes from yeast to plants to insects use RNA interference (RNAi) as a form of intracellular immune response against viral infection. Plants and animals have evolved diverse and extensive mechanisms of innate immunity. Some immune systems even operate at the colony level: in addition to their individual immune systems, social insects use smell to distinguish between members of the colony and potentially dangerous outsiders.

References

- [1] J. Maynard Smith and E. Szathmary. *The Major Transitions in Evolution*. Oxford University Press, Oxford, 1995.
- [2] E. Szathmary and J. Maynard Smith. The major evolutionary transitions. *Nature*, 374:227–232, 1995.
- [3] J. R. Krebs and R. Dawkins. *Animal signals: mind reading and manipulation*, chapter 15, pages 380–402. Blackwell Scientific, Oxford, 1984.
- [4] M. Lachmann, G. Sella, and E. Jablonka. On the advantages of information sharing. *Proceedings of the Royal Society of London B*, 267:1287–1293, 2000.
- [5] T. W. Pietsch and D. B. Grobecker. The compleat angler: Aggressive mimicry in an Antennariid anglerfish. *Science*, 201:369–370, 1978.
- [6] C. A. Munn. Birds that ‘cry wolf’. *Nature*, 319:143–145, 1986.

- [7] W. J. Dominey. Maintenance of female mimicry as a reproductive strategy in bluegill sunfish (*lepisomis macrochirus*). *Environmental Biology of Fishes*, 6:59–64, 1981.
- [8] M. D. Norman, J. Finn, and T. Tregenza. Dynamic mimicry in an Indo-Malayan octopus. *Proceedings of the Royal Society of London B*, 268:1755–1758, 2001.
- [9] James E. Lloyd. Aggressive mimicry in photuris: Firefly femmes fatales. *Science*, 149:653–654, 1965.
- [10] P. R. Y. Backwell, J. H. Christy, S. R. Telford, M. D. Jennions, and N. I. Passmore. Dishonest signalling in a fiddler crab. *Proceedings of the Royal Society of London B*, 267:719–724, 2000.
- [11] T. Akino, J. J. Knapp, J. A. Thomas, and G. W. Elmes. Chemical mimicry and host specificity in the butterfly *maculinea rebeli*, a social parasite of *myrmica* ant colonies. *Proceedings of the Royal Society of London B*, 266(1419-1426), 1999.
- [12] R. Steger and R. L. Caldwell. Intraspecific deception by bluffing — a defense strategy of stomatopods (Arthropoda, Crustacea). *Science*, 221:558–560, 1983.
- [13] J. Maynard Smith and D. G. C. Harper. Animal signals: Models and terminology. *Journal of Theoretical Biology*, 177:305–311, 1995.
- [14] M. Lachmann and C. T. Bergstrom. The disadvantage of combinatorial communication. *Proceedings of the Royal Society of London Series B*, 271:2337–2343, 2004.

- [15] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of eBay’s reputation system. In M. R. Baye, editor, *The Economics of the Internet and E-Commerce*. Elsevier Science, Amsterdam, 2002.
- [16] J. Boyd. In community we trust: Online security communication at ebay. *Journal of Computer-Mediated Communication*, 7(3), 2002.
- [17] eBay. Spoof Email Tutorial. <http://pages.ebay.com/education/spooftutorial/index.html>, 2007.
- [18] M. Spence. Job market signalling. *Quarterly Journal of Economics*, 87:355–374, 1973.
- [19] A. M. Spence. Time and communication in economic and social interaction. *Quarterly Journal of Economics*, pages 651–660, 1973.
- [20] A. Zahavi. Mate selection — a selection for a handicap. *Journal of Theoretical Biology*, 53:205–214, 1975.
- [21] A. Zahavi. The cost of honesty (Further remarks on the handicap principle). *Journal of Theoretical Biology*, 67:603–605, 1977.
- [22] G. W. F. Davis and P. O’Donald. Sexual selection for a handicap: A critical analysis of Zahavi’s model. *Journal of Theoretical Biology*, 57:345–354, 1976.
- [23] J. Maynard Smith. Sexual selection and the handicap principle. *Journal of Theoretical Biology*, 57:239–242, 1976.

- [24] M. Kirkpatrick. The handicap mechanism of sexual selection does not work. *American Naturalist*, 127:222–240, 1986.
- [25] N. Nur and O. Hasson. Phenotypic plasticity and the handicap principle. *Journal of Theoretical Biology*, 110:275–297, 1984.
- [26] A. Grafen. Biological signals as handicaps. *Journal of Theoretical Biology*, 144:517–546, 1990.
- [27] J. Maynard Smith. Honest signalling: The Philip Sidney game. *Animal Behaviour*, 42:1034–1035, 1991.
- [28] C. T. Bergstrom, Sz. Számadó, and M. Lachmann. Separating equilibria in continuous signalling games. *Philosophical Transactions of the Royal Society of London*, 357:1595–1606, 2002.
- [29] M. Lachmann, Sz. Számadó, and C. T. Bergstrom. Cost and conflict in animal signals and human language. *Proceedings of the National Academy of Sciences USA*, 98:13189–13194, 2001.
- [30] C. T. Bergstrom and M. Lachmann. Signalling among relatives. I. Is costly signalling too costly? *Philosophical Transactions of the Royal Society of London, Series B*, 352:609–617, 1997.
- [31] S. Rohwer. The social significance of avian winter plumage variability. *Evolution*, 29:593–610, 1975.
- [32] D. P. Whitfield. Plumage variability, status signalling and individual recognition in avian flocks. *Trends in Ecology and Evolution*, 2:13–18, 1987.

- [33] J. Maynard Smith and D. G. C. Harper. The evolution of aggression: Can selection generate variability? *Philosophical Transactions of the Royal Society of London, Series B*, 319:557–570, 1988.
- [34] S. Rohwer. Status signaling in harris sparrows: Some experiments in deception. *Behavior*, 61:107–129, 1975.
- [35] A. P. Moller. Social control of deception among status signalling house sparrows *Passer domesticus*. *Behavioral Ecology and Sociobiology*, 20:307–311, 1987.
- [36] P. Hammerstein, editor. *Genetic and Cultural Evolution of Cooperation*. MIT Press, Cambridge, MA, 2003.
- [37] C. T. Bergstrom, J. L. Bronstein, R. Bshary, R. C. Connor, M. Daly, S. A. Frank, H. Gintis, L. Keller, O. Leimar, R. Noë, and D. C. Queller. Interspecific mutualism: Puzzles and predictions regarding the emergence and maintenance of cooperation between species. In S. Bowles and P. Hammerstein, editors, *Dahlem Conference Report: Genetic and Cultural Evolution of Cooperation*. MIT Press, Cambridge, MA, 2003.
- [38] C. Anderson. The long tail. *Wired*, 12(10), 2004.
- [39] L. R. Gooding. Virus proteins that counteract host immune defences. *Cell*, 71:5–7, 1992.
- [40] D. M. Haig. Subversion and piracy: DNA viruses and immune invasion. *Research in Veterinary Science*, 70:205–219, 2001.

- [41] A. Liston and S. McColl. Subversion of the chemokine world by microbial pathogens. *BioEssays*, 25:478–488, 2003.
- [42] C. T. Bergstrom and R. Antia. How do adaptive immune systems control pathogens while avoiding autoimmunity? *Trends in Ecology and Evolution*, 2006. To appear January 2006.
- [43] R. C. Dorf and R. H. Bishop. *Modern Control Systems*. Prentice Hall, 10 edition, 2004.
- [44] H. Kitano. Computational systems biology. *Nature*, 420:206–210, 2002.
- [45] S. Kaech and R. Ahmed. Memory CD8+ T cell differentiation: initial antigen encounter triggers a developmental program in naive cells. *Nature Immunology*, 2:415–422, 2001.
- [46] R. Antia, C. T. Bergstrom, S. Pilyugin, S. M. Kaech, and R. Ahmed. Models of CD8+ responses. 1. What is the antigen-independent proliferation program? *Journal of Theoretical Biology*, 221:585–598, 2003.