Jumping to Bold Conclusions

A Review of *The Handicap Principle: A Missing Piece of Darwin's Puzzle*
by Amotz Zahavi and Avishag Zahavi
Oxford University Press: 1997. Pp. 286. £18.99, $30.00

Seth Bullock
Center for Adaptive Behavior and Cognition
Max Planck Institute for Human Development
Lentzeallee 94, D-14195 Berlin (-Dahlem)
Tel: 0049-30-82406–350, Fax: 0049-30-82406-399, Email: bullock@mpib-berlin.mpg.de

Occasionally, a grazing gazelle notices the approach of a predatory lion. Rather than immediately taking flight at top speed, she often jumps high in the air several times before fleeing. Perhaps these "stots" are warnings to close-by, possibly related, gazelles? If this is the case why is the warning so energetic? Surely evolution would favor less exhausting signals, since exhaustion is to be avoided when one is about to be pursued by a predator?

Amotz and Avishag Zahavi open their recent book with a different explanation for this behavior. They maintain that the stotting gazelle is not warning conspecifics of danger but informing the lion of its own escape ability. Because the lion has no desire to waste time and energy fruitlessly chasing uncatchable prey, this information is of use to it, but only if the information is truthful. It is here, the Zahavis claim, that an explanation is to be found for the otherwise inexplicable brio of the gazelle's display. The authors construe the stotting display as a handicap — a signaling behavior that incurs the depletion of the very quality it advertises. They interpret the gazelle's vigorous leaping to convey the honest message that she is fit enough and fast enough to waste just so much of her time and energy on jumping.

The "handicap principle," which underwrites the account of stotting offered above, has been the subject of energetic debate within the evolutionary biology literature since Israeli ornithologist Amotz Zahavi first presented it over 20 years ago (Zahavi, 1975, 1977). Despite initially attracting considerable skepticism, and enjoying only intermittent empirical and theoretical support over the next decade and a half, it has achieved increasing notoriety. More recently, the central tenets of the theory have earned acceptance as a result of the publication of a number of successful game-theoretic models (Grafen, 1990, being foremost among these). While the handicap principle has yet to gain the status of canonical edict (Zahavi having thus far, for instance, failed to garner much support for his contention that it eclipses "most" of Darwin's theory of sexual selection), both the vocabulary and explanatory perspective associated with it have attained a central position within current evolutionary thinking.

In addition to critiquing the Zahavis' new book, this review will give a flavor of the perspective afforded by the handicap principle and in doing so will highlight an area of biological thinking that might profit from the attention of the adaptive behavior community. But first, some understanding of the theoretical and methodological context into which the handicap principle was introduced is required.

Group Therapy

Following the mid-1960s decline of group selection as an acceptable explanatory strategy, many widespread and well-documented natural phenomena, which had previously been accounted for through some appeal to the worth of behavior at the level of social groups, were left drifting aimlessly without explanatory anchor. Topics as familiar as food sharing, flocking, and predator mobbing in birds, for instance, were temporarily rendered rudderless — loaded up with well-organized empirical data but lacking a theoretical helmsman. While several of these phenomena were quickly explained — the concepts of inclusive fitness and kin selection (Hamilton, 1964), for example, providing a unifying account of "altruism" among relatives that could explain the existence of sterile castes, allofeeding among relatives, etc. — signaling behaviors remained for some time the single largest and most embarrassingly unpiloted vessel on the high seas of evolutionary thought.

Mating displays, aggressive posturing, informative dances, begging cries, warning coloration, and danger signals, if honest, enable the efficient distribution of resources (food, sex, shelter, etc.). This efficiency derives from the flow of useful information between the members of an honest signaling system. Contrast a beehive, foraging as a unit on the basis of shared information, with the less efficient behavior of the same bees denied recourse to each others' signals.

However, although the increased efficiency afforded by honest signaling is of benefit to those groups that employ such signals, the individual producers and consumers of signals benefit from their membership in an honest signaling system only *on average*. While a group of individuals employing an honest signaling system might do better than a group of taciturn competitors, it is not clear that any individual member of an honest signaling system might not outperform its fellow group members through freeloading, bluffing, cheating, lying, double-crossing, exaggerating, misleading, or crying wolf.

If such malpractice reaps dividends (which in the case of highly related insects such as bees it may not) then it is plain that honesty within such a group is not the best policy. In such cases, the reasoning runs, surely natural signaling systems must collapse, undermined by the pernicious canker of falsity and deceit. Nevertheless, innumerable observations suggest that making and attending to signals is ubiquitous in nature. In addition to intercourse between birds, bees, and beasts, the cells within these creatures have also been demonstrated to traffic in information, and even the DNA *inside* these cells has been characterized as a genetic "code."

Zahavi's handicap principle was one of a number of attempts to reconcile the theoretically motivated suggestion that honest signaling should be evolutionarily unstable with the empirical evidence that signaling systems were the frequent product of evolution. What distanced the handicap principle from its competitors was its emphasis on honesty and cooperation rather than selfishness and competition, and what made Zahavi's suggestion particularly problematic was his assertion that *waste*, rather than economy, would provide an explanation for the existence and persistence of honest signaling in nature. On one front, Zahavi's handicap accounts, which were predicated upon a notion of extravagance, confronted group-theoretic accounts of signaling that had been underwritten by the well-established notion of evolution favoring efficiency. Simultaneously, the handicap principle challenged more recent individualistic accounts of communication, which regarded signaling systems as inherently unstable evolutionary arms races spiraling toward inevitable breakdown (Krebs & Dawkins, 1984), with the notion that telling the truth could be a reasonable evolutionary strategy. With hindsight, it seems inevitable that the theory would excite opposition.

The Economy of Waste

Zahavi's reasoning closely parallels (although was probably arrived at independently from) that of Thorstein Veblen, a turn-of-the-century sociologist who proposed the notion of "conspicuous consumption" (1899). For Veblen, the overindulgence exhibited by members of what he termed the "leisure class" could be understood as a demonstration of class membership; that is, the purchase of prohibitively expensive goods and services could be understood as an indicator of the procurer's wealth. This index of societal status was an effective one because those of lower status could not afford to make the "advertisements" of which wealthier individuals were capable. Indeed, at the lower extreme of the scale, the funds of the poorest individuals were more than accounted for by the demands of simply staying alive, leaving no extra money to "waste" upon "unproductive consumption."

Zahavi suggested that the utility of waste as an honest indicator was responsible for the form of all natural signals. As an initial line of support for this hypothesis Zahavi noted that many signaling systems involved what appeared to be needlessly costly signals. The peacock's tail, an image that was to become a potent symbol of the handicap principle and that graces the cover of the Zahavis' recent book, was highlighted as one such overly expensive signal. The long feathers that make up the male's tail serve no useful function other than attracting mates. Furthermore, they require considerable resources to develop, are hard to maintain, limit the male's ability to evade predators or resist parasites, and impose significant energetic costs in courtship, during which they are held erect and shaken vigorously.

Darwin (1871) had noted the seemingly unjustified extravagance of such mating displays and had been forced to explain it as resulting from the aesthetic sensibilities of the selecting sex. In contrast, Zahavi proposed that, far from satisfying the arbitrary coquettish whim of a choosy peahen, the displaying peacock was demonstrating his quality in a manner selected precisely for its capacity to reveal useful facts about his value as a prospective mate — by bearing a handicap. Zahavi employs the term handicap in much the same manner as it is used in certain sports — to denote a millstone that reduces performance by squandering some valuable resource. For racehorses, handicaps are additional **ballast**, for golfers, additional strokes, for archers, additional distance from the target. However, whereas these sporting handicaps are imposed to level the field, Zahavi pointed out that they also reveal the quality of the sportsmen. Thus, Zahavi's claim was that, although the specific form of natural signals was still essentially arbitrary (they might be visual, auditory, chemical, etc.), only those that satisfied handicap criteria would be honest, and thus stable.

The subsequent theoretical literature inspired and/or provoked by this thesis set about establishing whether such handicap criteria were logically possible, and if so what they might prove to be. After many years of disagreement among the authors of these models, the balance of opinion has shifted in favor of the basic premise of Zahavi's model — that the evolutionary stability of honest signals can be ensured by meeting conditions governing their cost. The details are however still at issue (see Johnstone, 1997, for an excellent recent review of the signaling literature).

With the recent publication of *The Handicap Principle: A Missing Piece of Darwin's Puzzle*, Zahavi has provided the first book-length effort toward demonstrating the extraordinary explanatory scope of the handicap principle. Zahavi's book is a family affair, coauthored with his wife Avishag (also an ornithologist), and translated from

Hebrew by his daughter and son-in-law. In it, the authors explore the application of handicap thinking to a catholic range of biological topics spanning intercellular signals, collaboration among social insects, the life-cycle of slime molds, and various human behaviors including suicide, dancing, and method acting.

This breakneck romp through a dizzying array of wildly differing subject matter is explicitly undertaken in an effort to convince the reader of the fundamental nature of the explanatory principles being employed — handicap thinking, the authors imply, is practically all-embracing. And, while individual jumps and twists in this story are frequently exciting and occasionally breathtaking, as with many roller-coaster rides, at times one wishes to get off. The primary motivation to **jump off early** stems from a suspicion that the ride is missing some structural supports. Zahavi has never had much time for formal models, and after devoting a paragraph at the outset of the book to note that his ideas have been validated by recent models, and that the nay-sayers have thus been dealt with, the authors proceed to write a book that is, on the one hand, admirably accessible and untroubled by complicated mathematics, but on the other, lacking in the rigor that typically accompanies such formal treatments.

As a result the book is in turn exasperating, challenging, engaging, and inspiring. Especially noteworthy are radical new perspectives offered on cooperation in social insect societies, and the altruism of slime molds (colonies of simple unrelated creatures within which only some specialized morphs are able to reproduce, but all strive to effect this reproduction), the concept of an extremely pervasive role for prestige within group-living creatures (which is intended to compete with kin-selectionist explanations), and the authoritative accounts of the social organization of Arabian babblers, a group-living species of bird that the Zahavis have been studying for the last 25 years.

This last aspect of the book stands out as its strongest. A clear and fascinating picture of the complex societal pressures within babbler communities is presented, and the aspects of their behavior that are understandable from the perspective of the handicap principle are among the book's most persuasive examples of handicap thinking. For instance, the roles of both allofeeding (birds feeding other birds) and acting as a group sentinel (watching for predators rather than, for instance, foraging) are explained from the perspective of maintaining prestige within the babbler community. Both activities are costly, in terms of time and wasted resources, yet both activities are performed by high-ranking birds. Furthermore, subordinate birds who attempt to carry out these activities (offering food to a superior, or attempting to replace a superior as sentinel) are often attacked and prevented from doing so by higher ranking birds. This competition for the "right" to be altruistic is inexplicable from the perspectives offered by competing accounts of such activities.

In contrast to the work on babblers, the final chapter, which deals with the possibility of explaining human behavior from the perspective of the handicap principle, epitomizes the problematic aspects of the book. Zahavi has used examples of human behavior throughout his writings on the handicap principle, sometimes apparently hypothetically or for some pedagogical or explicatory purpose. However, the arguments presented in the book tend to typify the kind of evolutionary account of human behavior that attracts the derision of nonevolutionary psychologists, anthropologists, sociologists, and the like.

Each argument follows the same format. Some piece of human morphology or behavior (from suicide, through sex and menstruation, to breasts and beards) is identified as being costly in some often only vaguely specified manner and is therefore considered to be a candidate for explanation from the perspective offered by the handicap principle. The Zahavis' worst such offense is perhaps their account of the evolutionary significance of necklaces and other such decorations, which amounts to little more than the claim that a "girl with a short neck cannot afford to wear a ribbon; such a decoration would make her appear ridiculously short-necked" (p. 216). What it is that is ridiculous about a short neck is not related. More importantly, the relationships between the cost of ridicule, the benefit of being perceived to be long-necked, and how the ability to bear such costs and exploit such benefits varies with the trait being advertised (neck length? mate quality? nutritional resources during development?) is left similarly unexpressed. Yet it is precisely the nature of these relationships that theoretical models have identified as crucial in determining whether handicap signaling may be evolutionarily stable and thus a strong candidate explanation for the continued existence of some natural signaling system.

This kind of informality also infects the Zahavis' more important ideas. At several points within the text they slide between different meanings of handicap, cost, signal, and so forth, eliding certain aspects, enhancing others, and neglecting conceptual differences.

For example, one discussion concerns the observation that it is impossible to make threatening or acquiescent noises without adopting certain postures. As the authors urge the reader to attempt to rise violently while letting out a relaxed sigh, or to remain comfortably seated, muscles slack, but simultaneously to make an aggressive roar, one is allowed a fanciful insight into the Zahavis' working life: after an exhausting day in the field, **indulging in** a spot of armchair experimentation with which to support a novel application of their favorite theory. However, to claim that the honest conveyance of threat or submission is stabilized by handicap costs is to risk confusing physical constraints on evolution with strategic ones. These two sorts of claim have been separated within the signaling literature (e.g., Maynard Smith & Harper, 1995) and deserve separate consideration. If the Zahavis' claim is that no creature can make a noise without revealing something of the state of its musculature then there is simply no room for dishonesty within this system. Talk of handicaps adds nothing to the notion that a certain cue simply cannot be faked. If instead the authors' claim is that although individuals might disguise their bodily state, the costs of doing so are too high, an account of what these costs might be, and how they might vary with the bodily state to be advertised, is required.

However, it would be unfair to censure the Zahavis too heavily on the issue of informality because many of their wilder assertions are proffered as bold hypotheses intended to provoke empirical studies, and many of the conceptual and theoretical problems that their book fails to address are yet to be adequately handled elsewhere within the evolutionary biology signaling literature. It is here that the adaptive behavior community may prove to be of service.

Individual Differences and Individual-Based Models

The issue central to the success of a handicap account is the relationship between the costs of the signaling behavior (in terms of energy, time, and risk), the benefits of this behavior (in terms of attaining some desired outcome, e.g., a copulation), and, importantly, how the manner in which fitness is influenced by these costs and benefits itself varies with the relevant individual differences being advertised (e.g., fighting ability, hunger, mate quality, etc.).

The current understanding of what is conveyed by the term "handicap criteria" can be captured by the claim that signaling will be evolutionarily stable when the net cost of signaling varies inversely with the magnitude of the trait being advertised (Bullock, 1997, 1998). In such cases, those signalers with more of the advertised trait will be able to signal more. It will be worth their while to signal more because in that way they will gain the valuable response they seek more often or to a greater extent. Bluffing or cheating will not be profitable, because amplifying one's advertisement will achieve, at best, a Pyrrhic victory — the increased benefits gained being outweighed by the increased costs suffered.

How likely is it that particular natural scenarios meet these criteria? How might net signaling cost vary with some important individual difference? The theoretical literature is littered with conflicting intuitions, and while the first empirical assessments of the costs and benefits of signaling behavior have begun to be published (see Kilner & Johnstone, 1997, for one recent review), such studies have so far failed to address clearly how the impact of these costs and benefits on signaler fitness varies with the relevant individual differences between signalers.

Consider a scenario in which an interloper makes a signal of aggressive intent to an observing harem holder. Grafen (1990) assumes that the fitness benefit enjoyed by an interloper that successfully routs its opponent *increases* with the quality of the interloper. Although this assumption is presented without support, presumably Grafen's reasoning is that a high-quality interloper will be more able to utilize a harem. For example, a stronger interloper will be able to successfully defend a harem for longer and, *ceteris paribus*, sire more offspring. In contrast, Adams and Mesterton-Gibbons (1995) suggest that in such situations, the benefit of attaining the retreat of an adversary might *decrease* with increasing interloper quality. They reason that "weak animals have more to gain by avoiding direct fights since they are less able to defend against injury" (p. 406). Similar disagreements frustrate attempts to model sexual advertisements, begging offspring, status badges, predator-prey signaling, and so forth.

The adaptive behavior community's focus upon the way in which high-level organizations arise from low-level interactions seems ideally suited to addressing these concerns. While game-theoretic models must explicitly stipulate the relationships between classes of strategists, evolutionary simulation models may employ what have been termed implicit fitness functions instantiated as processes that act over individual artificial entities (e.g., Wheeler & de Bourcier, 1995). The distinction between explicit and implicit fitness functions has been recognized for some time, but it has been made to do little work. Here, the advantage of models utilizing implicit fitness functions is that they may be constructed in such a manner as to capture the hypothesized nature of signaling interactions without committing the modeler to formalizing the manner in which fitness is determined by these interactions. As a result, the form of the fitness function implicit within the model may itself become the object of inquiry.

For example, Wheeler and de Bourcier (1995) model a scenario in which grazing animats move around a heterogeneous environment, signal their level of aggression to each other, and engage in agonistic interactions when they touch. Although the fitness consequences of interactions between multiple animats are not specified explicitly, the simulation allows the effects of population structure and density upon the success of different signaling strategies to be explored. This exploration can be understood as an analysis of the form of the fitness function implicit within the model's dynamics, although the authors do not describe their research in this way.

Questions such as "How do opportunity costs vary with sexual attractiveness?" or "How does the value of being overlooked by a predator vary with escape ability?" may be approached in a similar manner — as investigations into the structure of implicit fitness functions governing the evolution of various simple signaling simulations. The results of such inquiry will help to sharpen theoretical intuitions, and to guide and motivate empirical studies that attempt to address similar questions.

In addition to the possible benefits of implicit fitness functions for exploring the specifics of handicap signaling, evolutionary simulation models have many additional qualities to recommend them. Perhaps most importantly, they allow an exploration of a model's full evolutionary dynamics, in addition to merely characterizing the system's **equilibria** states. While theoretical biology has paid attention to the existence of honest signaling equilibria, it has very rarely addressed the attainability of such equilibria by populations evolving from some nonsignaling ancestral state. Simulation models are natural tools for such tasks (Bullock, 1997, 1998).

More generally, many competing evolutionary theories predict the same outcomes (they were, after all, arrived at in an attempt to account for the same empirical observations). However, these theories often differ in their hypotheses regarding the manner in which these outcomes are achieved. Such theories also differ in their predictions concerning from what initial conditions these outcomes may legitimately arise. As a result, attention to the transient evolutionary behavior of model populations, rather than their ultimate end states, may better help to highlight the differences between theories and hence decide between them.

Summary

The evolution of signaling systems has been of interest to the adaptive behavior community since its inception, yet the work resulting from this interest has consistently failed to address issues of current interest within evolutionary biology, typically preferring to demonstrate the potential for communicative behavior to evolve under some selective pressure favoring information exchange (e.g., MacLennan, 1991; Werner & Dyer, 1991). As a result this largely exploratory body of work has made little impact within the evolutionary biology literature. The time has come to demonstrate the worth of this modeling approach by producing research that interfaces with that published in biology journals and presented at biology meetings.

Current biological research into the handicap principle seems fertile ground upon which to establish such an interface because such research is raising issues that are difficult to address using the formal tools traditionally available to theoretical biologists. The Zahavis' new book should appeal to anyone interested in the evolution of signaling phenomena, cooperative behavior, group-living species, or altruism. More specifically, *The Handicap Principle* also provides an excellent source of innovative evolutionary ideas that demand formal treatments, empirical studies, and, most pertinent to this forum, appropriate evolutionary simulation models.

Despite the recent spate of successful game-theoretic treatments of the handicap principle, its worth as an explanation of the signaling behavior that permeates the natural world has yet to be demonstrated **conclusively**. While the validity of the handicap principle has repeatedly been questioned, the Zahavis**'** would no doubt claim that only a theory of genuine high quality would still be arousing the interest of biologists after nearly a quarter of a century spent shouldering the handicap of aggressive scrutiny exacerbated by informal presentation and compounded by theoretical controversy. Researchers in adaptive behavior can help to establish whether the bold hypotheses collected within the Zahavis' book are indeed missing pieces of Darwin's puzzle, or just ideas that, although interestingly shaped, are ultimately superfluous.

References

Adams, E. S., & Mesterton-Gibbons, M. (1995). The cost of threat displays and the stability of deceptive communication. *J. Theor. Biol.*, *175*, 405–421.

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands, P., & Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, pp. 454–463 Cambridge, MA. MIT Press.

Bullock, S. (1998). A continuous evolutionary simulation model of the attainability of honest signalling equilibria. In Adami, C., Belew, R. K., Kitano, H., & Taylor, C. E. (Eds.), *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life*, pp. 339–348 Cambridge, MA. MIT Press/Bradford Books.

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

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

Hamilton, W. D. (1964). The genetical evolution of social behaviour, I and II. *J. Theor. Biol.*, *7*, 1–16; 17–32.

Johnstone, R. A. (1997). The Evolution of Animal Signals. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (4th edition)., pp. 155–178. Blackwell, Oxford.

Kilner, R., & Johnstone, R. A. (1997). Begging the question: are offspring solicitation behaviours signals of need?. *Trends in Ecology and Evolution*, *12*(1), 11–15.

Krebs, J. R., & Dawkins, R. (1984). Animal Signals: Mind Reading and Manipulation. In Krebs, J. R., & Davies, N. B. (Eds.), *Behavioural Ecology: An Evolutionary Approach* (2nd edition)., pp. 380–402. Blackwell, Oxford.

MacLennan, B. (1991). Synthetic Ethology: An Approach to the Study of Communication. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 631–658 Redwood City, California. Addison-Wesley.

Maynard Smith, J., & Harper, D. G. C. (1995). Animal Signals: Models and Terminology. *J. Theor. Biol.*, *177*, 305–311.

Veblen, T. (1899). The theory of the leisure class. In Lerner, M. (Ed.), *The Portable Veblen*, pp. 53–214. Viking Press, New York. Collection published 1948.

Werner, G. M., & Dyer, M. G. (1991). Evolution of Communication in Artificial Organisms. In Langton, C. G., Taylor, C., Farmer, J. D., & Rasmussen, S. (Eds.), *Artificial Life II — SFI Studies in the Sciences of Complexity*, Vol. X, pp. 659–687 Redwood City, California. Addison-Wesley.

Wheeler, M., & de Bourcier, P. (1995). How not to murder your neighbour: Using Synthetic Behavioural Ecology to study aggressive signalling. *Adaptive Behavior*, *3*(3), 273–309.

Zahavi, A. (1975). Mate selection — A selection for a handicap. *J. Theor. Biol.*, *53*, 205–214.

Zahavi, A. (1977). The cost of honesty (further remarks on the handicap principle). *J. Theor. Biol.*, *67*, 603–605.