CHAPTER 25

The Evolution of Cognitive Bias

MARTIE G. HASELTON, DANIEL NETTLE, and PAUL W. ANDREWS

Humans, like other animals, see the world through the lens of evolved adaptations. In vision, for example, the experience of color is mediated by the adaptations of the eye, which in the human case uses wavelengths of electromagnetic radiation between about 380 and 760 nanometers, allowing us to see hues ranging from red to violet. But, there are other possible colors on earth. Recent work on bird species demonstrates that "blue" tits are actually ultraviolet (Hunt, Bennett, Cuthill, & Griffiths, 1998). The feathers of the male blue tit reflect ultraviolet radiation (300 to 400 nm), and females display a preference for males with the brightest ultraviolet crests (Hunt et al., 1998). Some reptiles, such as rattlesnakes, see light in the infrared range (see Goldsmith, 1990, for a review). Color is not an inherent property of an object; it is constructed by the interaction of reflected radiation in the environment with evolved visual mechanisms in the perceiver (Bennett, Cuthill, & Norris, 1994).

Using faculties of social perception, humans construct images of the social world in similar ways. Like color, sexual attractiveness is not a feature of the world that preexists the mechanisms that perceive it, and what is sexually attractive varies depending on the perceiver. Within humans, what appears attractive in a man depends on adaptively relevant variables that differ between female perceivers and within individual perceivers at different points in time. Women who are higher in physical attractiveness themselves find facially masculine men more attractive than do less attractive women (Little, Burt, Penton-Voak, & Perrett, 2001). Even more dramatic, women's ratings of men's attractiveness vary across the menstrual cycle, with more facially masculine men preferred near ovulation and less masculine men preferred at other times (Penton-Voak et al., 1999). Thus, a man that a woman sees as particularly attractive on one day might seem less so on another, even though he has not changed at all.

We are grateful to David Buss for generous editorial feedback and for pointing out the auditory looming phenomenon and the idea that outbreaks of disease evoke error management biases. Thanks also to Craig Fox and David Funder for helpful comments on an earlier draft of this chapter.

It appears that the "there" of familiar experience is one that the mind has a role in constructing. The mind translates the properties of the world, such as electromagnetic radiation and the contours of objects, into useful units of information. One of the most dramatic demonstrations of the role of the mind in our apprehension of the world is the existence of cognitive biases. A wide range of biases, which we review in the next section, has been discovered by psychologists. Where biases exists, individuals draw inferences or adopt beliefs where the evidence for doing so in a logically sound manner is either insufficient or absent.

As well as being interesting in their own right, biases are important to study because they often reveal the design of the mind. In this chapter, we present a three-category framework for understanding cognitive biases from an evolutionary perspective, and we discuss what biases in each category can tell us about the evolved mind. We conclude by describing the implications of this evolutionary psychological perspective on biases. For example, the functional specificity of these biases reveals the intricacy of the mind's design and supports the key hypothesis that the mechanisms of mind are domain-specific. The conclusion that many biases are not the result of constraints or mysterious irrationalities also speaks to the ongoing debate about human rationality. Our perspective suggests that biases often are not *design flaws,* but *design features.*

THE EVOLUTIONARY FOUNDATIONS OF COGNITIVE BIAS

> "Rational" decision-making methods . . . logic, mathematics, probability theory . . . are computationally weak: incapable of solving the natural adaptive problems our ancestors had to solve reliably in order to reproduce. . . . This poor performance on most natural problems is the primary reason why problem-solving specializations were favored by natural selection over general-purpose problem-solvers. Despite widespread claims to the contrary, the human mind is not worse than rational . . . but may often be better than rational.
>
> —Cosmides & Tooby, 1994, p. 329

Cognitive biases present something of a challenge to the evolutionary psychologist. Because they depart from standards of logic and accuracy, they appear to be design flaws instead of examples of good engineering. Cognitive traits can be evaluated according to any number of performance criteria (logical sufficiency, accuracy, speed of processing, etc.), and the value of a criterion depends on the question the scientist is asking. The question facing the evolutionary scientist is to identify whether a feature has been shaped by selection and, if so, to determine what its function is. Often the scientist has information about only what the feature does—its effects. However, selection often generates a tight fit between the design features and their effects. To the evolutionary psychologist, therefore, the evaluative task is not whether the cognitive feature is accurate or logical, but how well it solves a particular problem relative to other problems that it could potentially solve (i.e., whether the trait solves a problem with proficiency and specificity).

In evolutionary psychology, proficiency and specificity interact in the concept of *domain specificity.* As Tooby and Cosmides (1992) and others have argued, it is likely that the mind is equipped with function-specific mechanisms adapted for special purposes—mechanisms with special design for solving problems of mating, which are separate, at least in part, from those involved in solving problems of food choice, predator avoidance, and social exchange (see Boyer & Barrett, Chapter 3, this volume). In the evaluation of cognitive biases, demonstrating domain specificity in solving a particular problem is a part of building a case that the trait has been shaped by selection to perform that function. The evolved function of the eye, for instance, is to facilitate sight because it does this well (it exhibits proficiency), the features of the eye have the common and unique effect of facilitating sight (it exhibits specificity), and there are no plausible alternative hypotheses that account for the eye's features.

Some design features that appear to be flaws when viewed in one way are revealed to be adaptations when viewed differently. If we only consider the fact that high fevers make people feel miserable, and in extreme cases can lead to death, the capacity to develop fever appears a terrible flaw in design. However, if we ask what the evolved function of fever is, we come to learn that elevated body temperature may be a natural defense against pathogens. This hypothesis led to research showing that fever-reducing medicines increase susceptibility to infection and prolong its resolution, thus challenging the common use of aspirin as a treatment for upper respiratory infections and influenza (see Williams & Nesse, 1991, for a review). Viewed in light of evolution, the capacity for fever may in fact be well-designed.

In sum, there may be many evolutionary reasons for apparent design flaws, and a close examination often provides insight into the evolutionary forces that shaped them and their functions. We propose that analogous logic may be applied to understanding cognitive biases. Cognitive biases can arise for three reasons: (1) Selection may discover useful shortcuts that tend to work in most circumstances, though they fall short of some normative standards (heuristics); (2) biases can arise if biased solutions to adaptive problems resulted in lower error costs than unbiased ones (error management biases); and (3) apparent biases can arise if the task at hand is not one for which the mind is designed (artifacts). Table 25.1 presents this

Table 25.1
Evolutionary Taxonomy of Cognitive Biases

Type of Bias	Examples
Heuristic: Bias results from evolutionary or information processing constraints; mechanisms work well in most circumstances, but are prone to break down in systematic ways.	1. Use of stereotypes 2. Fundamental attribution "error" 3. One-reason decision strategies
Error management bias: Selection favored bias toward the less costly error; although error rates are increased, net costs are reduced.	1. Auditory looming (Figure 1) 2. Sexual overperception by men (Figure 2) 3. Commitment underperception by women (Figure 3) 4. Positive illusions
Artifact: Apparent biases and errors are artifacts of research strategies; they result from the application of inappropriate normative standards or placement of humans in unnatural settings.	1. Some instances of base-rate neglect in statistical prediction 2. Some instances of the confirmation bias

taxonomy. We do not intend these categories to be fully exhaustive or mutually exclusive; we do propose that they are a useful way of organizing research on cognitive bias and gaining insight into why biases occur.

HEURISTICS

Perhaps the most commonly invoked explanation for bias is that they are a by-product of processing limitations—because information processing time and ability are limited, humans must use shortcuts or rules of thumb that are prone to breakdown in systematic ways. This explanation for biases can be traced in large part to the influential work of Kahneman and Tversky (e.g., Tversky & Kahneman, 1974; see Gilovich, Griffin, & Kahneman, 2002, for a recent review; see Kahneman, 2003, for a recent theoretical treatment). Kahneman and Tversky demonstrated that human judgments often departed substantially from normative standards based on probability theory or simple logic. In judging the sequences of coin flips, for example, people assessed the sequence HTHTTH to be more likely than the sequence HHHTTT or HHHHTH. As Tversky and Kahneman (1974) pointed out, while in some sense representative, the first sequence contains too many alternations and too few runs. The "gambler's fallacy" is the expression of a similar intuition. The more bets lost, the more the gambler feels a win is now due, even though each new turn is independent of the last (Tversky & Kahneman, 1974).

Another example is the famous "Linda problem" (Tversky & Kahneman, 1983). Subjects read a personality description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and participated in antinuclear demonstrations." They were then asked to determine which of two options was more probable: (a) Linda is a bank teller or (b) Linda is a bank teller and active in the feminist movement. Although the conjunction cannot be more likely than either of its constituents, between 80% and 90% of subjects tend to select (b) as the more probable option. Tversky and Kahneman (1983) dubbed this effect the "conjunction fallacy."

Tversky and Kahneman attributed these and other biases to the operation of mental shortcuts: "People rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations" (1974, p. 1124). The gambler's fallacy and the conjunction fallacy are attributed to one of the most commonly invoked heuristics, *representativeness*, or the way in which A resembles or is representative of B. According to this account, alternating heads and tails is more representative of randomness than are series containing runs. The description of Linda is representative of a feminist; thus participants choose feminist and bank teller rather than bank teller alone.

Effects of Time and Motivation

The notion that biases result from the use of simplifying heuristics has logical appeal. As expressed by Arkes (1991), "The extra effort required to use a more sophisticated strategy is a cost that often outweighs the potential benefit of enhanced accuracy" (pp. 486–487). This cost can affect the evolution of cognitive mechanisms at two levels: (1) There may be costs in evolutionary terms because the development of certain brain circuits will either increase the length of ontogeny or remove potential energetic allocation away from the development of other mechanisms, and (2) there may be costs in real time because decisions using complex algorithms will take longer or require more attentional resources than decisions using simpler alternatives. Adaptive decisions often need to be made fast, and this may well constrain the type of strategies that are optimal. Evidence from a variety of sources demonstrates that people do indeed solve problems differently when under time pressure or when their motivations to be accurate are reduced.

Fiske (1993) proposed that the social perceptions of individuals occupying positions of higher power in social hierarchies are less accurate than those lower in the hierarchy. Those higher in power are more likely to endorse stereotypes about others than to attend to individuating information specific to the target being evaluated, which presumably enhances accuracy (Goodwin, Gubin, Fiske, & Yzerbyt, 2000). Individuals assigned more decision-making power in reviewing internship applications attend more to stereotype consistent information and less to stereotype disconfirming information (Goodwin et al., 2000). Similarly, in a study of two student groups competing for university funding, Ebenbach and Keltner (1998) found that individuals reporting more personal power judged their opponents' attitudes less accurately. A common interpretation of findings such as these is that lower power individuals occupy a more precarious social position and must, therefore, allocate more time and energy to social judgments; more powerful individuals enjoy the luxury of allocating their cognitive efforts elsewhere (Keltner, Gruenfeld, & Anderson, 2003).

Some of the best evidence for cognitive heuristics comes from the abundant literature on the *fundamental attribution error* (FAE). The FAE occurs when people are asked to make an inference about the mental state or underlying disposition of an actor. Although there are differences of opinion about the precise nature of the FAE (e.g., Sabini, Siepmann, & Stein, 2001), a relatively descriptive account is that people tend to infer that an actor's internal state corresponds to expressed behavior more than appears to be logically warranted by the situation (Andrews, 2001; Ross, 1977).

In the classic experiment, Jones and Harris (1967) gave subjects either a pro-Castro essay or an anti-Castro essay purportedly written by a student. In one set of variants (the choice variants), subjects were merely told that the student wrote the essay as part of a class. In the other variants (the no-choice variants), subjects were told that the professor assigned the student either the pro-Castro or the anti-Castro stance. In both variants, the corresponding inference is that the writer actually believes in the stance. In the no-choice variants, the situation suggests that the writer is less likely to believe in the stance than in the choice variants. Jones and Harris predicted that subjects would be agnostic about the writer's actual beliefs in the no-choice condition. However, subjects tended to make the corresponding inference even in the no-choice conditions—the so-called fundamental attribution error.

The FAE may result from a frugal cognitive heuristic that is usually effective. For example, personality does indeed exist, and when people make personality inferences, even about people with whom they have had only brief interactions, they have some predictive validity (e.g., Colvin & Funder, 1991). If the FAE is the result of a frugal cognitive heuristic, one prediction is that it should take cognitive effort to avoid it. People who avoid the FAE do take longer on the task (Yost & Weary, 1996), and they are more likely to commit the FAE under conditions of cognitive load (Trope & Alfieri, 1997). A second prediction is that incentives for more refined judgments should diminish reliance on the heuristic. Indeed, people are less likely to make the FAE when they are told they will be held accountable for their attributions (Tetlock, 1985) or when they are given a monetary incentive for making correct attributions (Vonk, 1999). Moreover, people in more interdependent societies (e.g., China or Korea) are less likely to make the FAE than people in the United States or Europe (Norenzayan, Choi, & Nisbett, 2002). A possible explanation is that greater interdependency makes it more important to make accurate attributions of others' mental states and internal dispositions (Neuberg & Fiske, 1987).

Effectiveness of Heuristics

Heuristics and biases research has generated much debate, for example, about the ecological validity of the problems, the conclusions to be drawn about the adequacy of everyday decision making, and the precision of the heuristics proposed (Gigerenzer, 1996, 2000; Hertwig & Gigerenzer, 1999; Kahneman & Tversky, 1996; see also Biases as Artifacts section). From an evolutionary psychology perspective, though, what is of interest is the broad consensus that human decision making relies on a repertoire of simple, fast, heuristic decision rules to be used in specific situations. Much experimental work has focused on the cases where these rules lead to illogicality or error, but the assumption is that over a broad range of fitness-relevant past scenarios, they were highly effective.

The direct demonstration of the effectiveness of heuristic decision rules has only just begun. Gigerenzer and Goldstein (1996) showed that a family of simple decision-making rules that uses only one datum can work as well or better than more complex algorithms that use all available information, for example, the *recognition heuristic.* When asked to make judgments about which of two alternatives will be higher on some criterion variable (e.g., who will succeed in a sports contest or which city is larger), someone who uses the heuristic will choose the alternative that is most familiar. For example, when asked which city has a larger population, San Diego or San Antonio, German students tend to guess right: San Diego (Goldstein & Gigerenzer, 1999). Paradoxically, American students tend to get it wrong. This is the *less-is-more effect*—American students cannot use recognition because both cities are known, so they rely on other cues, which are often invalid.

Gigerenzer and Goldstein (1996) augmented these surprising results with formal simulations that pitted one-reason heuristics, in this case, the *take-the-best* algorithm, against computationally sophisticated algorithms suggested by other cognitive scientists. *Take-the-best* assumes a decision tree structure. It starts with recognition. If recognition is a predictor of the criterion variable and one item in the decision task is recognized but the other is not, the recognized option is selected. If recognition does not apply (e.g., both options are recognized), then you move to the next step in the tree, searching memory for the most valid cue that discriminates between the alternatives. If the cue has a positive value for one alternative but not for the other, *take-the-best* is completed. If both have a positive cue value, the next cue is retrieved from memory, and so on. The algorithm is fast because it is relatively simple, and it is frugal because it looks up only as much information as it needs. And, surprisingly, it performs as well or better than classically optimal algorithms that use all of the information available to the decision maker (e.g., multiple regression).

These results offer an existence proof. Decision-making adaptations can be simple but still as effective as complex strategies on real-world tasks. If researchers in the laboratory can exploit reliable features of the informational environment to create simple but highly effective reasoning strategies, natural selection can do so as well. Because selection has shaped different decision strategies for different adaptive problems, it seems unlikely that there is a single, general *take-the-best* or *recognition* adaptation. Rather, these simple strategies, and others like them, form the armamentarium that natural selection has tended to use in creating decision-making adaptations. We propose that combinations of these strategies are used by an array of distinct, domain-specific, evolved mental mechanisms. (For two interesting examples in the mate selection context, see Dugatkin, 1996; Miller & Todd, 1998.)

In sum, there is ample evidence of cognitive bias and error in humans. Some of these biases may result from the use of shortcuts, which are often effective. There is also evidence that relaxing constraints or increasing motivation for accuracy can improve reasoning in some domains. For these effects, it is important to note that the constraints explanation is not complete. Why are these biases the defaults? We have suggested that dispositional inference may be the default because personality has predictive power (also see Haselton & Buss, 2003). In the following section, we further suggest that the direction and content of biases is not arbitrary. Although these biases might arise in part because of the mind's limited computational power, the particular forms they assume serve the fitness interests of the perceiver (also see Kenrick & Maner, in press; Krebs & Denton, 1997).

ERROR MANAGEMENT BIASES

> Laboratory research on "error" . . . attracts much attention because it appears to have dismal implications for social reasoning. These implications are illusory, however, because an error is not the same thing as a "mistake."
>
> —Funder (1987)

Error Management Theory

Error management theory (EMT; Haselton & Buss, 2000; Haselton & Nettle, 2004) applies the principles of signal detection theory (Green & Swets, 1966) to judgment tasks to make predictions about evolved cognitive design. The central idea is that any cognitive mechanism can produce two types of error: a false positive (adopting a belief that is false) and a false negative (failing to adopt a belief that is true). The equivalents couched in perceptual terms are detecting a stimulus that does not exist (false positive) and failing to detect a stimulus that is real (false negative), but the logic is the same in either formulation.

On the face of it, it would seem that an optimal mechanism would make no errors of either type. However, many real-world tasks of judgment are probabilistic and, therefore, include an irreducible amount of uncertainty. Auditory judgment, for example, is rendered uncertain by the presence of ambient noise, and some error is likely to occur however good the mechanism.

Crucially, the consequences for the organism of making the two types of error may not be the same. Fleeing from an area that contains no predator may be inconvenient but is much less costly than the failure to detect a predator that really is close by. EMT predicts that an optimal decision rule will minimize not the crude rate of error, but the net effect of error on fitness. Where one error is consistently more damaging to fitness than the other, EMT predicts that a bias toward making the less costly error will evolve; it is better to make *more* errors overall as long as they are of the relatively cheap kind. The magnitude and direction of bias are predicted to be affected by two factors: the asymmetry of the cost of the two errors (the bias will be toward making the less costly error, and larger asymmetries produce larger biases) and the amount of uncertainty in the task (biases are expected only when judgments are uncertain). For mathematical formalism of this logic and the expectations of EMT, see Haselton and Nettle (2004).

We have argued that many apparent biases may reflect the operation of mechanisms designed to make inexpensive, frequent errors rather than occasional disastrous ones (Haselton & Nettle, 2004). Table 25.2 on pages 732–733 provides examples by outlining the domain in which the effect occurs, the hypothesized costs of errors, and the expected outcome. In the following subsections, we discuss each of the entries in Table 25.2 along with other illustrative effects (for a complete review and analysis, see Haselton & Nettle, 2004).

Protective Biases

Broadly speaking, the possible error management effects we have identified fall into three somewhat overlapping clusters. The first are physically protective biases.

Auditory Looming A perceptual example is *auditory looming* (Neuhoff, 2001). People judge a sound that is rising in intensity to be closer, and approaching more rapidly, than an equidistant sound that is falling in intensity. In a series of experiments involving speakers moving on cables, Neuhoff (2001) shows that auditory looming leads to biased perceptions of the proximity of moving sound sources, as well as a general tendency to underestimate the distance of sound sources. Subjects judge an approaching sound source to be closer by than a receding one, when in fact they are located at distances equally far away from the subject (see Figure 25.1 on p. 734). There is a clear error management interpretation of this effect: It is better to be ready for an approaching object too early than too late (Neuhoff, 2001).

Allergy, Cough, and Anxiety Nesse (2001) argued for the *smoke detector principle* in bodily systems designed to protect from harm. He describes medical examples such as allergy and cough where a protective system is often mobilized in the absence of real threat. These defense systems appear to be overresponsive. Dampening them with drugs or treatment actually results in few troublesome effects on the recipient (Nesse, 2001). Psychological defense mechanisms such as anxiety are also easily evoked, especially in connection with things likely to have been dangerous in the ancestral environment, such as spiders, snakes, and potentially

Table 25.2
A Selection of Adaptive Biases

Category and Domain	False Positive (FP)	Costs of FP	False Negative (FN)	Costs of FN	Result
Protective: Approaching sounds	Ready too early	Low	Struck by source	High	Bias toward under-estimating time to arrival
Protective: Foodstuffs	Eat a food type that is safe	Low	Ingest toxin or pathogen	High	Bias toward acquiring permanent aversion on the basis of one piece of evidence of toxicity
Protective: Diseased persons	Avoid noninfectious person	May be low, depending on relationship	Become infected	Often very high	Tendency to avoid persons with physical afflictions, even if noninfectious
Social: Men's inference of female sexual interest	Inferring sexual interest where there is none	Rejection—relatively low	Inferring no interest when there is interest	Missed reproductive opportunity—high	Sexual overperception by men
Social: Women's inference of commitment	Inferring interest to commit where there is none	Desertion—high	Inferring unwillingness to commit where there is willingness	Delayed start to reproduction—relatively low	Underperception of commitment by women.

Social:					
Social exchange	Attempt to free ride and get caught	Potential ostracism, especially in collectivist social situations—high	Cooperate when one could free ride	Give up an unnecessary benefit in exchange—relatively low	Bias toward cooperation
Self and Future:					
Beliefs about future achievements	Believe you can achieve things when you cannot	Low (if costs of failure are low)	Believe you cannot achieve things when you could	High (if benefit of success is high)	Optimistic bias (where benefits of success exceed costs of failue)
Self and Future:					
Control of events	Believe you control events that you have no control over	Low (if control behavior is inexpensive)	Believe your behavior ineffective when it is effective	Leads to passivity; potentially high	Illusion of control, superstitions

Adapted from *The Evolution of the Paranoid Optimist: An Integrative Model of Cognitive Biases,* by M. G. Haselton and D. Nettle, 2004, Manuscript under review, University of California, Los Angeles.

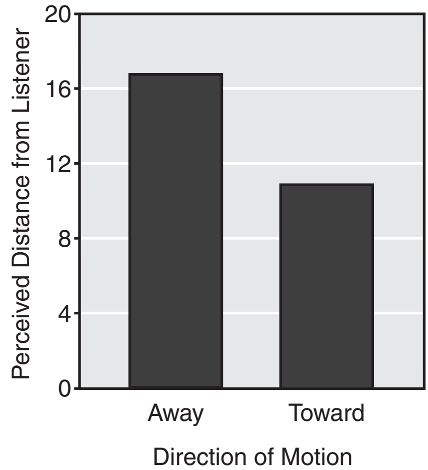

Figure 25.1 Auditory Looming. Subjects estimated the distance of sound sources that were moving away from them and toward them. The sources were of equal intensity and were, on average, equally far away. The true mean distance was 20 feet, and thus there was a tendency toward underestimation in general as well as a specific bias for approaching sound sources. *Source:* From "An Adaptive Bias in the Perception of Looming Auditory Motion," by J. G. Neuhoff, 2001, *Ecological Psychology, 13,* pp. 87–110.

dangerous persons (Mineka, 1992; Seligman, 1971; Tomarken, Mineka, & Cook, 1989). A tendency for anxiety mechanisms to produce false positives is a plausible explanation for the observed prevalence of phobias and anxiety disorders (Nesse, 2001).

Food Aversions Food aversions may be similarly biased. Lasting aversion to a food is reliably acquired, in humans and other species, following a single incidence of sickness after ingestion (Garcia, Hankins, & Rusiniak, 1976; Rozin & Kalat, 1971). Given one data point (sickness followed the food type on one occasion), the system treats the food as if it is always illness inducing. There are again two possible errors here: The false positive may be inconvenient, but the false negative is more likely to be fatal. The system appears biased toward overresponsiveness to avoid illness.

Aversion to Diseased or Injured Persons Similar logic predicts an aversion to the ill. Little evidence of illness or contamination is required to provoke avoidance of a person, whereas much stronger evidence is required to warrant the inference that someone is safe or free from disease (Kurzban & Leary, 2001; Park, Faulkner, & Schaller, 2003). The error management account is similar to that for food aversions: The false negative (failing to avoid someone with a contagious disease) is highly costly, whereas the false positive (avoiding contact with a noncontagious person) may be inconvenient but is unlikely to be injurious. Thus, disease-avoidance mechanisms will be biased and tend to evince disgust and avoidance at many stimuli that are safe (e.g., deformity as a result of injury rather than disease). Such a bias may well be involved in the panics associated with outbreaks of diseases such as SARS and Mad Cow disease, when more mundane risks nearer to home may be far greater objective dangers.

Biases in Interpersonal Perception

The second cluster of biases concerns interpersonal perception. Some of these apply to both men and women, whereas others are sex-specific. EMT predicts differential effects of sex in those domains where the costs and benefits of an outcome differ reliably between the sexes—domains that include, most obviously, mating and its consequences.

Sexual Overperception Courtship communications are often ambiguous. Does a smile convey mere friendliness, for example, or does it mean more? Haselton and Buss (2000) proposed that men possess a bias in interpreting cues to a woman's sexual interest. For ancestral men, they argued, it would have been more costly in reproductive currency to miss a woman's sexual interest than to overestimate it. In the evolutionary past, men's reproduction was limited primarily by the number of women of reproductive age to whom they were able to gain sexual access (Buss, 1994; Symons, 1979; Trivers, 1972). Men who were more often successful in mating with greater numbers of women tended to outreproduce their fellow male competitors. For women, partner number played a smaller role in reproduction. Because of their relatively heavy investment in each offspring produced and long interbirth interval, finding partners with heritable quality and a strong disposition to invest would likely have had a larger impact on reproductive success than would securing additional mating opportunities (Buss, 1994; Symons, 1979; Trivers, 1972). Thus, for men but not women, a missed mating opportunity with a fertile partner because of underestimated sexual interest would have been a high cost fitness error. An overestimation error may have carried some costs (e.g., to reputation), but these costs would have been lower overall. Several sources of evidence support the sexual overperception hypothesis (see Haselton & Buss, 2000, for a review). In initial meetings between male and female strangers, for example, men tend to rate women's flirtatiousness and sexual interest higher than do women (Abbey, 1982). The difference in male and female ratings is obtained when men's ratings are compared to the target woman's self-ratings and when compared to third-party women's ratings of the target woman's sexual interest (Abbey, 1982; Haselton & Buss, 2000). Similar results are obtained in naturalistic studies. Haselton (2003) asked women and men to report past instances of sexual misperception. Women reported more instances in the past year in which men overestimated their sexual interest than in which men underestimated it, suggesting a male sexual overperception bias. Men reported roughly equal numbers of overperception and underperception errors on the part of women, suggesting no bias in women (see Figure 25.2 on p. 736). Recently, Maner and colleagues documented further evidence of sexual overperception in men (Maner et al., in press). They induced romantic arousal, fear, or a neutral emotion state by showing films to study participants. They then asked participants to examine photographs of faces with neutral facial expressions for hidden cues ("micro-expressions") to their actual emotion state. Relative to the other film conditions, in the romantic arousal condition, men

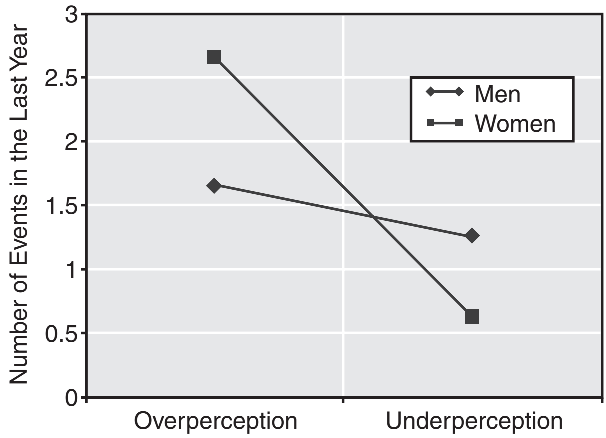

Figure 25.2 Sexual Overperception by Men: Women ($n = 102$) and Men ($n = 114$) Reported Past Experiences in Which a Member of the Opposite Sex Erroneously Inferred Their Sexual Interest. Within the last year, women reported significantly more overperception errors committed by men than underperception errors ($p < .001$), suggesting that men systematically overestimate sexual intent. Men reported roughly equal numbers of overperception and underperception errors committed by women, suggesting no bias in women's sexual inferences ($p > .05$). *Source:* From "The Sexual Overperception Bias: Evidence of a Systematic Bias in Men from a Survey of Naturally Occurring Events," by M. G. Haselton, 2003, *Journal of Research in Personality, 37,* pp. 43–47. Used with permission of Elsevier.

increased their attribution of sexual interest for female faces. They showed no such effect for male faces. After viewing the romantic film, women did not increase attribution of sexual interest for either male or female faces. Conceptually similar sexual overperception effects are observed in the behaviors of males in some bird, insect, and mammalian species (Alcock, 1993, chap. 13; Domjan, Huber-McDonald, & Holloway, 1992).

Commitment Skepticism The reverse asymmetry may have applied to ancestral women as they decoded men's courtship communications surrounding commitment (Haselton & Buss, 2000). For a woman, inferring long-term commitment interest in a man in whom it was absent could have resulted in postconceptive abandonment, a high cost error associated with lowered offspring survival (Hurtado & Hill, 1992). Underestimating a man's commitment could also result in nontrivial costs such as delaying reproduction, but these costs, Haselton and Buss (2000) hypothesized, would have been lower on average than costs associated with desertion. Women may, therefore, possess a bias toward underestimating men's interest in commitment. Women do indeed rate the level of commitment communicated by male courtship behaviors such as giving of gifts and verbal affirmations of love as lower than do men (Haselton & Buss, 2000). In contrast, women and men tend to agree on the level of commitment communicated by women on the basis of the same behaviors (Haselton & Buss, 2000). In vignette studies, women more than men infer deceptive intentions in a man who is conveying his interest in forming a long-term relationship with a woman he would like to take out on a date (Andrews, 2002; see Figure 25.3).

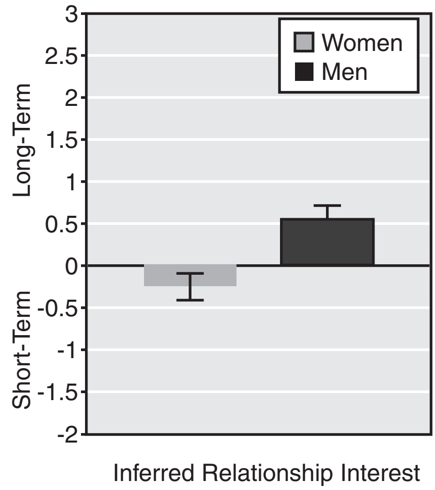

Figure 25.3 Commitment Underperception by Women. Women (n = 108) and men (n = 60) were asked to evaluate a potentially deceptive scenario. The man in the scenario might be interested in a short-term affair, although he claims to be interested in a long-term relationship. Participants' task was to infer the likelihood that the target man was deceptively interested in a short-term relationship or honestly conveying interest in a long-term relationship. Women (lighter colored bar) inferred greater deception (interest in a short-term relationship only; $p < .01$) than did men. *Source:* From *Attributing Honesty to a Signal Purporting to Reveal Mental State* by P. W. Andrews, June 2002, paper presented at the Human Behavior and Evolution Society Conference, Rutgers, NJ.

Negative Outgroup Stereotypes Humans appear to possess a bias toward inferring that members of competing coalitions (or out-groups) are less generous and kind (Brewer, 1979) and more dangerous and mean (Quillian & Pager, 2001) than are members of their own group. This may be understood as an adaptive bias. For ancestral humans, the costs of falsely assuming peacefulness on the part of an aggressor were likely to outweigh the comparatively low costs of elevated vigilance toward aggression, especially for inferences regarding out-group members. For in-group members, elevated inferences of aggressiveness would have carried the additional costs of within-coalition conflict; hence the negative bias might be expected to be small or nonexistent for in-group members. Schaller and colleagues proposed that cues signaling increased risk of injury, such as ambient darkness, might increase these effects because they raise the costs of failures to detect aggression and protect the self (Schaller, Park, & Mueller, 2003). As predicted, subjects who completed a rating task in a darkened laboratory increased their endorsement of racial and ethnic stereotypes connoting violence, relative to those who participated in a brightly lit room (Schaller et al., 2003). Darkness had no effect on other negative stereotypes of out-group others (e.g., laziness or ignorance; Schaller et al., 2003).

Social Exchange Bias Behavioral economists are puzzled by the fact that people cooperate in economic games with incentive structures favoring defection (Camerer & Thaler, 1995; Caporael, Dawes, Orbell, & van de Kragt, 1989; Henrich et al., 2001; Sally, 1995). In the one-shot prisoner's dilemma game, for example, participants are expected to defect rather than to cooperate. If partner A cooperates while B defects, partner A suffers a greater loss than if he or she had defected. The interaction is not repeated, so there is no incentive to signal cooperativeness, nor is there prior information about reputation that might serve to provide clues about the partner's cooperative disposition. Yet, cooperation often occurs, as it does in other one-shot tasks.

Yamagishi and colleagues hypothesized that cooperation in one-shot games results from the operation of a social exchange bias that manages the costs of errors in social exchange (Yamagishi, Terai, Kiyonari, & Kanazawa, 2003). They propose that the costs of falsely believing one partner can defect without negative social consequences are often higher than cooperating when he or she could safely defect. This asymmetry holds when the costs of "unneeded" cooperation are relatively low (e.g., a low dollar amount is lost) or when the social costs of failing to cooperate (potential ostracism) are high. The costs of ostracism may be particularly high in interdependent social contexts, in which cooperation is either highly valued or especially necessary (Yamagishi, Jin, & Kiyonari, 1999). In Japanese collectivist samples where exchanges are relatively closed to outsiders, cooperation in one-shot experiments is indeed higher than in the more individualist U.S. samples (Yamagishi et al., 1999). Also as predicted, when participants are led to think of the game as an exchange relationship (by making forecasts about their exchange partner's behavior), they cooperate more than when they are not (Yamagishi et al., 2003; see also Savitsky, Epley, & Gillovich, 2001; and Williams, Case, & Govan, 2003; for related predictions).

Note that this bias can be conceptualized as some combination of error management, as in the Yamagishi account, and an artifact of modern living because in an ancestral environment the probability of reencountering individuals would have been high and social reputation effects very potent. Thus, people may be predisposed to expect negative consequences of nonprosocial behavior even when, objectively, such consequences are unlikely to follow. Note, too, that the bias toward prosociality is the subject of competing explanations, which take quite different explanatory stances (Bowles & Gintis, 2002; Gintis, Bowles, Boyd, & Fehr, 2003; Henrich & Boyd, 2001; Price, Cosmides, & Tooby, 2002), and it is unexplored whether these are complementary or competing accounts to the social exchange bias.

Biases in Self-Judgment

The third cluster of biases concerns judgment about the self and personal efficacy. For a complete review, see Haselton and Nettle (2004). Here we briefly discuss the representative example of the *positive illusions.*

Positive Illusions Positive illusions are a well-known cluster of findings in judgment tasks concerning the self (Taylor & Brown, 1988). Individuals display unrealistically positive perceptions of their own qualities (Alicke, 1985), their likelihood of achieving positive outcomes in the future (Weinstein, 1980), and their degree of control over processes in the environment (Alloy & Abramson, 1979; Rudski, 2000). Two classes of evolutionary explanation have been proposed for such tendencies. First, individuals may have been selected to optimize the impression of their qualities that they display to observers. Given that observers will not be able to accurately assess such qualities directly, individuals may display behaviors that strategically enhance the qualities conveyed (Sedikides, Gaertner, & Toguchi, 2003).

An alternative explanation is in error management terms. Nettle (2004) outlines such an explanation, building on the interpretation of the positive illusions given by Taylor and Brown (1988). In evaluating a possible behavior, there are two possible errors. We may judge that the behavior is worthwhile when in fact it achieves nothing to promote fitness, or we may judge that a behavior is not worthwhile when in fact it would have enhanced fitness to do it. The former error (a false positive) leads to behaviors that are useless, whereas the latter (a false negative) leads to passivity. The costs of the false positive and false negative errors may not be symmetrical—that is, trying and failing may not matter very much, whereas failing to try could be very costly, at least relative to competitors. Thus, evolution can be expected to produce mechanisms biased toward positive illusion in domains where there is uncertainty about outcomes, and the cost of trying and failing is reliably less than that of not trying where success was possible (Nettle, 2004). Note that this account does not predict blanket optimism, but optimism where fitness gains are potentially high relative to the cost of passivity.

The self-enhancement and error management accounts are not mutually exclusive, and it has not been possible to demonstrate their relative importance in producing positive illusions. It may be possible to have them make differential predictions. Take a scenario with some chance of a moderate gain and some chance of extreme physical pain, for example. Impression management would seem to predict an intuition of optimism about the chances of success because it is based on strategic presentation of desirable qualities such as courage and robustness. Error management, however, seems to predict an intuition of pessimism because it is designed to avoid very costly errors under uncertainty. Performance on such tasks could be significantly affected by both sex of respondent and audience presence. Such possibilities await empirical investigation. (For a fuller account of the error management approach and its predictions, see Haselton & Nettle, 2004.)

BIASES AS ARTIFACTS

One criticism of classic heuristics and biases research is that the strategies for identifying bias and evaluating cognitive performance might not be appropriate. Similarly, if problems presented in the laboratory are not those for which the human mind is designed, it should not be surprising that their responses appear to be systematically irrational. In this section, we discuss two general categories of artifact effects: evolutionarily invalid problem formats and evolutionarily invalid problem content.

Problem Formats

Gigerenzer (1997) proposed that tasks intended to assess human statistical prediction should present information in frequency form. Natural frequencies, such as the number of times an event has occurred in a given time period, are more readily observable in nature. Probabilities (in the sense of a number between 0 and 1) are mathematical abstractions beyond sensory input data, and information about the base rates of occurrence is lost when probabilities are computed (Cosmides & Tooby, 1996). Bayesian calculations involving frequencies are, therefore, computationally simpler than equivalent calculations involving probabilities, relative frequencies, or percentages. Whereas probability calculations need to reintroduce information about base rates, frequency calculations do not because this part of the computation is already done within the frequency representation itself (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2001).

Humans should possess the ability to estimate the likelihood of events given certain cues. If this skill is a part of human reasoning, however, tasks involving probability input are less likely to reveal it than are tasks involving natural frequencies. Indeed, frequency formats do improve performance in tasks like the Linda problem. Whereas a probability format produces violations of the conjunction rule in between 50% and 90% of subjects, frequency formats decrease the rate of error to between 0 and 25% (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983; also see Cosmides & Tooby, 1996). The frequency interpretation of these results, however, is controversial (e.g., Gigerenzer, 1996; Kahneman & Tversky, 1996). Attempts to rule out competing hypotheses about confounds have generally supported the frequency hypothesis (see Cosmides & Tooby, 1996, experiments 5 and 6; also see Hertwig & Gigerenzer, 1999, experiment 4), but neither perspective appears to perfectly account for all of the available data (see Mellers, Hertwig, & Kahneman, 2001).

A related set of questions has been raised about the conversational pragmatics (see Grice, 1975) of bias-eliciting word problems. Hertwig and Gigerenzer (1999) note that *probability* (or *probable*) in the Linda task is a polysemous term with both mathematical and nonmathematical interpretations. Participants who are asked whether it is more probable that Linda is a *bank teller* or a *feminist bank teller* could infer that the researcher is asking which is a better description of Linda, in which case the conjunction effect is not technically an error. If participants assume that researchers are following maxims of conversational pragmatics, which would lead them to assume that all information provided by the researchers is *relevant* to solving the problem, a mathematical interpretation of the word *probability* is less likely, because it renders the description of Linda's personal commitments irrelevant (Hertwig & Gigerenzer, 1999).

Problem Content

The perspective on cognitive design we have described suggests that researchers should not necessarily expect good performance in tasks involving abstract rules of logic. Falsification-based logic is sufficiently difficult for humans that university courses in logic, statistics, and research design attempt to teach it to students (often with only mixed success). Students have to learn that, in scientific practice, to test hypotheses they must look not only for confirmatory evidence but also for potentially falsifying evidence. If only confirmatory evidence is found, but no falsifying evidence, the hypothesis is supported although it still could be wrong. If even one piece of unambiguous falsifying evidence is found, the hypothesis is contradicted and is very unlikely to be right. It is difficult for students to intuitively grasp that searching for falsifying evidence is a stronger test of the hypothesis than confirmatory evidence.

Wason (1983) empirically confirmed this in the laboratory using a task that required subjects to determine whether a conditional rule (if *p* then *q*) had been broken. He demonstrated that subjects recognized that confirmatory evidence (the presence of *p*) was relevant to the decision, but they often failed to check for falsifications of the rule (the absence of *q*). Research using the Wason task revealed a variety of apparent content effects (Johnson-Laird, Legrenzi, & Legrenzi, 1972; Wason & Shapiro, 1971), in which subjects' performance dramatically changed for the better.

In a series of now-classic experiments, Cosmides (1989) demonstrated that a number of the content effects could be attributed to a cheater-detection algorithm. When the content of the conditional rule involves social exchange (if you take the benefit [*p*], then you pay the cost [*q*]), people are spontaneously induced to look not only for benefits taken (*p*) but also costs not paid (not *q*), and performance dramatically increases from 25% correct (Wason, 1983) to 75% correct (Cosmides, 1989; see Cosmides & Tooby, Chapter 20, this volume, for an extensive discussion).

The conclusion to be drawn from these studies is not that humans are good at using abstract rules of logic. Rather, it is that humans have evolved problem-solving mechanisms tailored to problems recurrently present over evolutionary history. When problems are framed in ways congruent with these adaptive problems (e.g., social contract violation), humans can be shown to use appropriate reasoning strategies. The rarity of falsificatory choices in the nonsocial versions of the Wason task may reflect not so much *error* as the fact that the mental schemata tapped into by the problem are those for updating beliefs about probabilistic associations in the environment, which is not a deontic task but an indicative one and thus requires not falsification logic but Bayesian updating. When the question is construed in this sense, subjects' choices on the nonsocial versions are close to a Bayesian optimum (Oaksford & Chater, 1994).

In summary, many documented bias effects could reflect the application of normative standards that are not entirely appropriate for evaluating human performance. The content of problems also has been shown to have a strong effect on the approach that subjects take to reasoning; thus a normative standard that is abstract and content blind is bound to find human performance aberrant.

CONCLUSIONS

For most of its history, research on cognitive and social bias has been dominated by the failure and bleak implications of heuristics (see Kruger & Funder, in press). In a foundational paper in the heuristics and biases approach, Kahnemann and Tversky (1973) stated that "[people] rely on a limited number of heuristics which sometimes yield reasonable judgments and sometimes lead to severe and systematic errors" (p. 237). This relatively tempered viewpoint became exaggerated over the years. A *Newsweek* magazine account of the heuristics and biases research summarized it as showing that "most people . . . are woefully muddled information processors who often stumble on ill-chosen short-cuts to reach bad conclusions" (cited in Gigerenzer, Todd, & The ABC Research Group, 1999, p. 27). In reflecting on the history of social psychology, Aronson (1999) noted that "odious behavior 'sin' is at the heart of [the] most powerful research in social psychology" (p. 104). Browsing journals in social psychology, behavioral economics, and social cognition reveals a proliferation of seemingly foolish bias effects (see Kruger & Funder, in press).

Adopting an evolutionary perspective turns this focus on its head. Natural selection is the force responsible for creating the intricate designs with an improbably perfect match to their environments. Complex visual systems with specialized features tailored to species' differing ecologies have evolved several times independently (Goldsmith, 1990). Reproductive adaptations allow animals to reproduce small copies of themselves, developmentally intact, complete with miniature versions of the adaptations that will enable their own reproduction. And, natural selection is responsible for the most complex system known, the human brain. How could natural selection produce systems that equip the brain that are prone to fail as a rule and succeed only in exceptional cases?

The conceptual tide might now be turning. There has been a recent shift toward artifactual and adaptive explanations for bias, as well as a demonstration that simple mechanisms (heuristics) can function well in their proper domains. This reconceptualization has stimulated new developments in psychological theory and empirical research. Documenting content effects in biases—where bias effects emerge, recede, or reverse depending on the content of the judgment at hand—suggests that the mind does contain computationally distinct mechanisms governing reasoning in functionally distinct domains. Results demonstrating the presence of adaptive biases where they might logically be expected in one sex but not in the other and protective biases in response to stimuli that were ancestrally dangerous (but their conspicuous absence in response to modern threats) are key pieces of evidence in the debate about domain specificity. On the empirical side, these newer breeds of explanation cannot reasonably be dismissed as *just-so* stories. Although controversy about their interpretation remains, researchers from divergent perspectives have tested competing predictions about classic effects and contributed their findings to the body of knowledge in psychology. The adaptive bias explanation we have featured in this chapter, error management theory, has also stimulated investigation on particular biases that were predicted a priori (e.g., women's commitment skepticism).

These new developments do not necessarily diminish the lessons learned from earlier research. We occupy a world that is governed by novel economic rules, and knowledge of the ways in which our evolved psychology causes us to behave in ways that contrast with our self-interest in light of these rules should prove substantively important to human happiness (e.g., Thaler & Bernartzi, 2004). However, the recent amendments to theory do suggest a substantial overhaul to the conclusion that human judgment is fundamentally flawed, at least in the ways in which it has been depicted over the past three decades. When we observe humans in adaptively relevant environments, we can observe impressive design of human judgment that is free of irrational biases. Because of trade-offs in error costs, true biases also prove more functional than we would think based on first intuition. Some genuine biases might be functional features designed by the wisdom of natural selection.

REFERENCES

Abbey, A. (1982). Sex differences in attributions for friendly behavior: Do males misperceive females' friendliness? *Journal of Personality and Social Psychology, 42,* 830–838.

Alcock, J. (1993). *Animal behavior: An evolutionary approach* (5th ed.). Sunderland, MA: Basic Books.

Alicke, M. D. (1985). Global self-evaluation as defined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49,* 1621–1630.

Alloy, L. B., & Abramson, L. Y. (1979). Judgment of contingency in depressed and non-depressed subjects: Sadder but wiser? *Journal of Experimental Psychology: General, 108,* 443–479.

Andrews, P. W. (2001). The psychology of social chess and the evolution of attribution mechanisms: Explaining the fundamental attribution error. *Evolution and Human Behavior, 22,* 11–29.

Andrews, P. W. (2002, June). *Attributing honesty to a signal purporting to reveal mental state.* Paper presented at the Human Behavior and Evolution Society Conference, Rutgers, NJ.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications for debiasing. *Psychological Bulletin, 110,* 486–498.

Aronson, E. (1999). Adventures in social psychology: Roots, branches, and sticky new leaves. In A. Rodrigues & O. V. Levine (Eds.), *Reflections on 100 years of social psychology.* New York: Basic Books.

Bennett, A. T. D., Cuthill, I. C., & Norris, K. J. (1994). Sexual selection and the mismeasure of color. *American Naturalist, 144,* 848–860.

Bowles, S., & Gintis, H. (2002). Homo reciprocans. *Nature, 415,* 125–128.

Brewer, M. B. (1979). Ingroup bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86,* 307–324.

Buss, D. M. (1994). *The evolution of desire: Strategies of human mating.* New York: Basic Books.

Camerer, C., & Thaler, R. (1995). Ultimatums, dictators and manners. *Journal of Economic Perspectives, 9,* 337–356.

Caporael, L., Dawes, R. M., Orbell, J. M., & van de Kragt, A. J. (1989). Selfishness examined. *Behavioral and Brain Sciences, 12,* 683–739.

Colvin, C., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology, 60,* 884–894.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition, 31,* 187–276.

Cosmides, L., & Tooby, J. (1994). Better than rational: Evolutionary psychology and the invisible hand. *American Economic Review, 84,* 327–332.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1–73.

Domjan, M., Huber-McDonald, M., & Holloway, K. S. (1992). Conditioning copulatory behavior to an artificial object: Efficacy of stimulus fading. *Animal Learning and Behavior, 20,* 350–362.

Dugatkin, L. A. (1996). Interface between culturally based preferences and genetic preferences: Female mate choice in Poecilia reticulata. *Proceedings of the National Academy of Sciences, USA, 93,* 2770–2773.

Ebenbach, D. H., & Keltner, D. (1998). Power, emotion, and judgmental accuracy in social conflict: Motivating the cognitive miser. *Basic and Applied Social Psychology, 20,* 7–21.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123–129.

Fiske, S. T. (1993). Controlling other people: The impact of power on stereotyping. *American Psychologist, 48,* 621–628.

Funder, D. C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin, 101,* 75–90.

Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1976). Flavor aversion studies. *Science, 192,* 265–266.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review, 103,* 592–596.

Gigerenzer, G. (1997). Ecological intelligence: An adaptation for frequencies. *Psychologische Beitrage, 39,* 107–129.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world.* New York: Oxford University Press.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge.

Gintis, H., Bowles, S., Boyd, R., & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior, 24,* 153–172.

Goldsmith, T. H. (1990). Optimization, constraint and history in the evolution of eyes. *Quarterly Review of Biology, 65,* 281–322.

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.

Goodwin, S. A., Gubin, A. A., Fiske, S. T., & Yzerbyt, V. T. (2000). Power can bias subordinates by default and by design. *Group Processes and Intergroup Relations, 3,* 227–256.

Green, D. M., & Swets, J. A. (1966). *Signal detection and psychophysics.* New York: Wiley.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics, III: Speech acts* (pp. 41–58). New York: Academic Press.

Haselton, M. G. (2003). The sexual overperception bias: Evidence of a systematic bias in men from a survey of naturally occurring events. *Journal of Research in Personality, 37,* 43–47.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78,* 81–91.

Haselton, M. G., & Buss, D. M. (2003). Biases in social judgment: Design flaws or design features. In J. Forgas, K. Williams, & B. von Hippel (Eds.), *Responding to the social world: Implicit and explicit processes in social judgments and decisions.* New York: Cambridge.

Haselton, M. G., & Nettle, D. (2004). *The paranoid optimist: An integrative evolutionary model of cognitive biases.* Manuscript under review, Los Angeles, University of California.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208,* 79–89.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., et al. (2001). Cooperation, reciprocity and punishment in 15 small-scale societies. *American Economic Review, 91,* 73–78.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making, 12,* 275–305.

Hoffrage, U., Lindsey, S., Hertwig, R., & Gigerenzer, G. (2001, May 4th). Statistics: What seems natural (response to Butterworth). *Science, 292,* 855.

Hunt, S., Bennett, A. T. D., Cuthill, I. C., & Griffiths, R. (1998). Blue tits are ultraviolet tits. *Proceedings of the Royal Society of London, Series B, Biological Sciences, 265,* 451–455.

Hurtado, A. M., & Hill, K. R. (1992). Paternal effect on offspring survivorship among Ache and Hiwi hunter-gatherers. In B. S. Hewlett (Ed.), *Father-child relations: Cultural and biosocial contexts* (pp. 31–55). New York: Aldine De Gruyter.

Johnson-Laird, P., Legrenzi, P., & Legrenzi, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63,* 495–500.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3,* 1–24.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist, 58,* 697–720.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237–251.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103,* 582–591.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003). Power, approach, and inhibition. *Psychological Review, 110,* 265–284.

Kenrick, D. T., & Maner, J. (in press). One path to balance and order in social psychology: An evolutionary perspective. *Behavioral and Brain Sciences.*

Krebs, D. L., & Denton, K. (1997). Social illusions and self-deception: The evolution of biases in person perception. In J. A. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 21–47). Hillsdale, NJ: Erlbaum.

Krueger, J. I., & Funder, D. C. (in press). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences.*

Kurzban, R., & Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 123,* 187–208.

Little, A. C., Burt, D. M., Penton-Voak, I. S., & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society of London, Series B, Biological Sciences, 268,* 39–44.

Maner, J. K., Kenrick, D. T., Becker, V., Robertson, T. E., Hofer, B., Neuberg, S. L., et al. (in press). Functional projection: How fundamental social motives can bias interpersonal perception. *Journal of Personality and Social Psychology.*

Mellers, B., Hertwig, R., & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12,* 269–275.

Miller, G. F., & Todd, P. M. (1998). Mate choice turns cognitive. *Trends in Cognitive Sciences, 2,* 190–198.

Mineka, S. (1992). Evolutionary memories, emotional processing, and the emotional disorders. *Psychology of Learning and Motivation, 28,* 161–206.

Nesse, R. M. (2001). The smoke detector principle: Natural selection and the regulation of defenses. *Annals of the New York Academy of Sciences, 935,* 75–85.

Nettle, D. (2004). Adaptive illusions: Optimism, control and human rationality. In D. Evans & P. Cruse (Eds.), *Emotion, evolution and rationality* (pp. 191–206). Oxford, England: Oxford University Press.

Neuberg, S. L., & Fiske, S. T. (1987). Motivational influences on impression formation: Outcome dependency, accuracy-driven attention, and individuating processes. *Journal of Personality and Social Psychology, 53,* 431–444.

Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13,* 87–110.

Norenzayan, A., Choi, I., & Nisbett, R. E. (2002). Cultural similarities and differences in social inference: Evidence from behavioral predictions and lay theories of behavior. *Personality and Social Psychology Bulletin, 28,* 109–120.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101,* 608–631.

Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with disabilities. *Journal of Nonverbal Behavior, 27,* 65–87.

Penton-Voak, I. S., Perrett, D. I., Castles, D., Burt, M., Koyabashi, T., & Murray, L. K. (1999). Female preference for male faces changes cyclically. *Nature, 399,* 741–742.

Price, M., Cosmides, L., & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological adaptation. *Evolution and Human Behavior, 23,* 203–231.

Quillian, L., & Pager, D. (2001). Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime. *American Journal of Sociology, 107,* 717–767.

Ross, L. (1997). The intuitive psychologist and his shortcomings: Distortions in attribution processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 174–214). New York: Academic Press.

Rozin, P., & Kalat, J. W. (1971). Specific hungers and poison avoidances as adaptive specializations of learning. *Psychological Review, 78,* 459–486.

Rudski, J. M. (2000). Illusion of control relative to chance outcomes. *Psychological Reports, 87,* 85–92.

Sabini, J., Siepmann, M., & Stein, J. (2001). The really fundamental attribution error in social psychological research. *Psychological Inquiry, 12,* 1–15.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and Society, 7,* 58–92.

Savitsky, K., Epley, N., & Gilovich, T. (2001). Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. *Journal of Personality and Social Psychology, 81,* 44–56.

Schaller, M., Park, J. H., & Mueller, A. (2003). Fear of the dark: Interactive effects of beliefs about danger and ambient darkness on ethnic stereotypes. *Personality and Social Psychology Bulletin, 29,* 637–649.

Sedikides, C., Gaertner, L., & Toguchi, Y. (2003). Pancultural self-enhancement. *Journal of Personality and Social Psychology, 84,* 60–79.

Seligman, M. E. P. (1971). Phobias and preparedness. *Behavior Therapy, 2,* 307–320.

Symons, D. (1979). *The evolution of human sexuality.* New York: Oxford University Press.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193–201.

Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

Thaler, R. H., & Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of Political Economy, 112*(1, Pt. 2), S164–S187.

Tomarken, A. J., Mineka, S., & Cook, M. (1989). Fear-relevant selective associations and covariation bias. *Journal of Abnormal Psychology, 98,* 381–394.

Tooby, J., & Cosmides, L. (1992). The psychological foundations of culture. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 19–136). New York: Oxford University Press.

Trivers, R. L. (1972). Parental investment and sexual selection. In B. Campbell (Ed.), *Sexual selection and the descent of man, 1871–1971* (pp. 136–179). Chicago: Aldine Publishing Company.

Trope, Y., & Alfieri, T. (1997). Effortfulness and flexibility of dispositional judgment processes. *Journal of Personality and Social Psychology, 73,* 662–674.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1121–1131.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293–315.

Vonk, R. (1999). Effects of outcome dependency on correspondence bias. *Personality and Social Psychology Bulletin, 25,* 382–389.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. Evans (Ed.), *Thinking and reasoning: Psychological approaches.* London: Routeledge and Kegan Paul.

Wason, P. C., & Shapiro, D. (1971). Natural and contrived experience in a reasoning problem. *Quarterly Journal of Experimental Psychology, 23,* 63–71.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39,* 806–820.

Williams, G. W., & Nesse, R. M. (1991). The dawn of Darwinian medicine. *Quarterly Review of Biology, 66,* 1–22.

Williams, K. D., Case, T. I., & Govan, C. L. (2003). Impact of ostracism on social judgments and decisions: Explicit and implicit responses. In J. Forgas, K. Williams, & W. von Hippel (Eds.), *Responding to the social world: Implicit and explicit processes in social judgments and decisions.* New York: Cambridge University Press.

Yamagishi, T., Jin, N., & Kiyonari, T. (1999). Bounded generalized reciprocity: Ingroup favoritism and ingroup boasting. *Advances in Group Processes, 16,* 161–197.

Yamagishi, T., Terai, S., Kiyonari, T., & Kanazawa, S. (2003). *The social exchange heuristic: Managing errors in social exchange.* Manuscript under review, Center for Advanced Study in the Behavioral Sciences, Stanford, CA.

Yost, J. H., & Weary, G. (1996). Depression and the correspondent inference bias: Evidence for more effortful cognitive processing. *Personality and Social Psychology Bulletin, 22,* 192–200.