

The Evolution of Cognitive Bias

Martie G. Haselton

University of California, Los Angeles

Daniel Nettle

University of Newcastle upon Tyne

Damian R. Murray

University of California, Los Angeles

Chapter to appear in D. M. Buss *The Evolutionary Psychology Handbook*, 2nd Edition, Wiley.

Draft: December 12, 2014

Word count (text and references) = 9,971

The Evolution of Cognitive Bias

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)

On the surface, cognitive biases appear to be somewhat puzzling when viewed through an evolutionary lens. 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, and so on. The value of a criterion depends on the question the scientist is asking. To the evolutionary psychologist, however, the evaluative task is not whether the cognitive feature is accurate or logical, but rather how well it solves a particular problem, and how solving this problem contributed to fitness ancestrally. Viewed in this way, if a cognitive bias positively impacted fitness it is not a design *flaw* – it is a design *feature*. This chapter discusses the many biases that are probably not the result of mere constraints on the design of the mind or other mysterious irrationalities, but rather are adaptations that can be studied and better understood from an evolutionary perspective.

By cognitive bias, we mean cases in which human cognition reliably produces representations that are systematically distorted compared to some aspect of objective reality. We note that the term bias is used in the literature in a number of different ways (see, e.g., Johnson et al., 2013; Marshall et al., 2013; Nettle & Bateson, 2012). We do not seek to make commitments about these definitions here; rather, we use bias throughout this chapter in the relatively noncommittal sense defined above.

Foundations of Cognitive Bias

An evolutionary psychological perspective predicts that the mind is equipped with function-specific mechanisms adapted for special purposes—mechanisms with special design for

solving problems such as mating, which are separate, at least in part, from those involved in solving problems of food choice, predator avoidance, and social exchange (e.g. Kenrick, Neuberg, Griskevicius, Becker, & Schaller, 2010). 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 one were to only consider the idea that selection favors the maximization of direct reproductive success, for example, the fact that human females lose reproductive capability many years before death would appear a design flaw. However, there is evidence that women in traditional societies can enhance their inclusive fitness by transferring investment to their daughters' daughters as soon as the latter are of reproductive age (Volland & Beise, 2002). Viewed in this light, female menopause might be very well-designed (Hawkes 2003).

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. Analogous logic may be applied to understanding cognitive biases. We argue that cognitive biases can arise for at least three reasons. First, selection may favor useful short-cuts that tend to work in most circumstances, though they fall short of some normative standards (*heuristics*); second, apparent biases can arise if the task at hand is not one for which the mind is designed (*artifacts*); and, third, biases can arise if biased response patterns to adaptive problems

resulted in lower error costs than unbiased response patterns (*error management biases*). As well as being interesting in their own right, the investigation of biases offers the capacity to reveal the contours of the evolved mind by revealing the problems it appears to have been designed to solve: whereas “accurate” perceptions do little to constrain hypotheses about cognitive design, discovering bias can often reveal it.

Since the original edition of this Handbook, the volume of work investigating error management biases has grown rapidly. Therefore, we discuss heuristics and artifacts only briefly and focus on newer work on error management biases (for a more detailed evolutionary discussion of heuristics and artifacts see Haselton et al., 2009). We do not intend the three categories of bias to be fully exhaustive or mutually exclusive; we offer them instead as a useful way of organizing research on cognitive bias and gaining insight into why biases exist.

Heuristics

Perhaps the most commonly invoked explanation for bias is as a necessary 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 break-down in systematic ways. Kahneman and Tversky (1973) 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 HHHHHTH. As Tversky and Kahneman (1974) pointed out, while in some sense representative, the first type of sequence is improbable—it 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).

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 are more representative of randomness than are series containing runs.

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. There may be costs in evolutionary terms, since 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. There may also be costs in real time, since decisions using complex algorithms will often 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.

One example of the effects of motivation is the fact that the social perceptions of individuals occupying positions of higher power in social hierarchies are often less accurate than those lower in the hierarchy (Fiske, 1993). 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-inconsistent information (Goodwin et al, 2000). Similarly, in a study of two student groups competing for university funding, individuals reporting more personal power judged their opponents' attitudes less accurately (Ebenbach and Keltner, 1998). A common interpretation of findings such as these is that lower power individuals occupy a more precarious social position and they must therefore allocate more time and energy to social judgments; more powerful individuals enjoy the luxury of allocating their cognitive efforts elsewhere (Galinsky, Magee, Inesi, & Gruenfeld, 2006; Keltner, Gruenfeld, & Anderson, 2003).

Overall, there is ample evidence of cognitive bias and error in humans. Some of these biases may result from the use of short-cuts, which are often effective. For these effects, however, it is important to note that a "processing limitations" explanation is not complete. Of all possible equally economical cognitive short-cuts, why were these particular ones favored by selection? In the *error management biases* section below we suggest that the direction and content of biases is not arbitrary. Selection has sculpted the ways that limited computational power is deployed so as best to serve the fitness interests of humans over evolutionary time.

Biases as Artifacts

One criticism of classic heuristics and biases research (e.g., Tversky & Kahneman, 1974) is that the strategies for identifying bias and evaluating cognitive performance might not be fully appropriate. If problems presented in the laboratory *are not those for which the human mind is designed*, we should not be surprised that people's responses appear to be systematically irrational.

One type of artifact arises from *evolutionarily novel problem formats*. Gigerenzer (1997) proposed that tasks intended to assess human statistical prediction should present information in frequency (rather than probability) format, given that natural frequencies, such as the number of times an event has occurred in a given time period, are more readily observable in nature. In contrast, 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 since this part of the computation is already “done” within the frequency representation itself (Hoffrage, Lindsey, Hertwig, & Gigerenzer, 2001).

According to this perspective, humans will 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 famous “Linda problem.” Whereas a probability format produces violations of the conjunction rule in between 50 and 90% of respondents, frequency formats decrease the rate of error to between 0 and 25% (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983; but see Mellers, Hertwig, & Kahneman, 2001). More recent research suggests that probability formats pose serious problems for medical doctors: three quarters of doctors surveyed misinterpreted the meaning and application of “survival rates,” and journals frequently publish papers in which these probability statistics are misused in interpreting results (Gigerenzer and Wegwarth, 2013).

A second artifact can arise from *evolutionarily novel 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 (with only mixed success). 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 (Wason and Shapiro 1971; Johnson-Laird, Legrenzi and Legrenzi 1972), in which 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; also see Cosmides, Barrett, & Tooby, 2010, for a more recent update that replicates these findings and helps to rule out alternative explanations proposed by critics).

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 (such as social contract violation), humans can be shown to use appropriate reasoning strategies.

Error Management Biases

Error Management Theory

Error management theory (EMT; Haselton & Buss 2000; Haselton & Nettle, 2006; Johnson et al., 2013) applies the principles of signal detection theory (Green & Swets 1966) to judgment tasks, in order to make predictions about evolved cognitive design. An error management framework views cognitive mechanisms not so much as “truth seekers,” (as has been previously thought, e.g. Fodor, 2001), but as adaptation executors (e.g. Tooby & Cosmides, 1990). The central tenet of this framework is that cognitive mechanisms can generally produce two types of errors: false positives (taking an action that would have been better not to take), and false negatives (failing to take an action that would have been better to take).

An optimal mechanism would make no errors of either type. However, most real world judgment tasks are probabilistic and 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 fitness costs of making each type of error are rarely equal. Fleeing from an area that contains no predator results in a small inconvenience cost, but is much less costly than the failure to flee from a predator that really is close by. EMT predicts that an optimal decision rule will minimize not the total error rate, 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 towards making the less costly error will evolve – this is because it is better to make *more* errors overall as long as they are relatively cheap. Overall, then, EMT predicts that biases will evolve in human judgments and evaluations that fit all of the following criteria: 1) they involve some degree of noise or uncertainty, 2) they have consequences for fitness and reproductive success, and 3) they are consistently associated with asymmetrical costs (where more asymmetry leads to larger

biases). For mathematical formalism of this logic and the expectations of EMT, see Haselton and Nettle (2006) and Johnson et al. (2013).

Within this framework, many ostensible faults in human judgment and evaluation may reflect the operation of mechanisms designed to make inexpensive, frequent errors rather than occasional disastrous ones (Haselton & Nettle, 2006; Johnson et al., 2013). In the decade since the publication of the first edition of this volume, the scope of EMT research has expanded, with streams of research documenting functionally biased judgments across a variety of fitness-relevant domains. In this section we highlight key examples across these domains (for reviews containing additional examples, see Haselton & Galperin, 2013; Haselton et al., 2009; Haselton and Nettle, 2006; Johnson et al., 2013).

Error management biases can be generally sorted into three broad categories: biases pertaining to judgments of threat, biases pertaining to evaluations of interpersonal relationships, and biases pertaining to evaluations of the self (following Haselton & Nettle, 2006). Table 2 provides examples within each of these categories, the hypothesized costs of each type of error in a given domain, and the expected outcome for each.

Threat-Relevant Biases

Several biases might protect people from threats to physical safety or to health. We begin by discussing the former.

Auditory Looming. People tend to 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, people showed biased perceptions of the proximity of moving sound sources, as well as a general tendency to underestimate the distance of sound sources (Neuhoff, 2001). People judged an approaching sound source to be closer by

than a receding one, when in fact the sounds are located at distances equally far away. 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).

Recent work has shown that individuals in poorer physical condition—measured by both heart rate recovery time and physical strength—have larger auditory looming biases than individuals in better physical condition (Neuhoff, Long, and Worthington, 2012). An error management interpretation of this relationship is that individuals with reduced motor capacity require a larger “margin of safety.” In another recent study, people exposed to an infant cry showed larger auditory looming biases than people not exposed to the cry. And, conversely, people exposed to an infant *laugh* showed smaller auditory looming biases than people not exposed to the laugh (Neuhoff, Hamilton, Gittelsohn, & Mejia, 2014). Female participants showed larger shifts in auditory looming biases in response to these infant stimuli—a pattern also found in a follow-up study (Neuhoff et al., 2014). Because infants required direct care from mothers more so than from fathers (e.g., due to breastfeeding), these effects suggest that self-protective biases like auditory looming are tuned to threats associated with care for vulnerable offspring. At a more general level, adaptively patterned variation in auditory looming demonstrates that error management biases are not fixed, but are responsive to cues of variation in threat.

Movement of Threatening Objects. Might there be analogous phenomena in the perception of visual threats? In one recent study, people judged the speed of an approaching spider, ladybug, or rubber ball (Witt and Sugovic, 2013). Although all objects moved at the same speed, spiders were judged to be moving more quickly than the other objects. Further, when people were given the task of “blocking” the spider, they judged the spider as approaching them

faster when they used a smaller paddle relative to a larger paddle, demonstrating that the bias was enhanced when avoiding the spider was more difficult.

Properties of Physical Landscapes. There are asymmetric costs of injury from underestimating the height of a cliff, and perhaps erroneously judging it safe to jump, than from overestimating it and finding a different means of navigation. Consistent with this idea, people tend to judge the height of a vertical surface as greater when looking from the top rather than the bottom (Jackson & Cormack, 2007; Stefanucci & Proffitt, 2009).

A similar example involves the perceived steepness of hills. In one series of studies, people consistently overestimated the steepness of hills—both real and computer-simulated (Proffitt, Bhalla, Gossweiler, & Midgett, 1995). Failing to properly descend a steep hill is far more costly than failing to properly ascend it. An error management perspective therefore predicts that this bias towards overestimating slopes will be greater when people view from the top than from the bottom, which was exactly what is found (Proffitt et al., 1995). Making the situation even more precarious increases the bias even more – people standing on skateboards at the top of hills perceive greater steepness than those standing flat on the earth (Stefanucci, Proffitt, Clore, & Parekh, 2008).

Food Aversions. Lasting aversion to a food is reliably acquired, in humans and other species, following a single incidence of sickness after ingestion (Rozin & Kalat 1971; Garcia et al. 1976). Given one data point (sickness following 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 individuals who have superficial cues that might connote the threat of infectious disease. 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 non-contagious person) may have small social or interpersonal costs, but is unlikely to have significant negative fitness consequences. Given the fact that infectious disease has represented one of the key selective forces throughout human history (e.g. Inhorn & Brown, 1990) disease avoidance mechanisms are expected to be especially biased and tend to evince avoidance of many stimuli that are in fact safe.

This significant bias towards false positives in assessing cues of disease threat has far-reaching social and societal implications, and may lie at the root of many forms of stigmatization and prejudice, including racism, ageism, homophobia, and anti-fat prejudice (e.g. Kurzban & Leary, 2001). Hypersensitivity to disease threat leads to stigmatization or avoidance of individuals who pose no risk of disease transmission whatsoever, yet display cues that were associated with disease threat ancestrally. Individuals with noninfectious morphological anomalies, such as prominent birthmarks, activate avoidant responses (Zebrowitz & Montpare, 2006). Similarly, individuals with clearly noninfectious physical disabilities are also implicitly associated with disease (Park, Schaller, & Faulkner, 2003), as are obese individuals (Park, Schaller, & Crandall, 2007). Individuals displaying symptoms of HIV/AIDS are also implicitly associated with the threat of infectious disease, despite knowledge that this ailment is not infectious through superficial contact. These individuals are frequently regarded as disgusting (e.g. Herek, 1999) and they, along with their families, are often ostracized from their communities (Gerbert, Sumser, & Maguire, 1991). Other patently noninfectious afflictions that

result in social distancing include cancer (Greene & Banerjee, 2006) and physical disfigurements (Houston & Bull, 1994).

The strength of these implicit associations is predicted by the extent to which individuals perceive themselves to be vulnerable to infectious disease. Individuals who tend to be more worried about disease threat have stronger implicit associations between infectious disease and both obesity and physical disability, and also have more negative attitudes towards obese and physically disabled individuals (Lieberman, Tybur, & Latner, 2012; Park et al., 2003, 2007). Moreover, making a threat of disease temporarily salient amplifies these prejudicial cognitions (Park et al., 2003, 2007). Other evidence suggests that prejudicial cognitions regarding elderly people are greater among people who feel more chronically vulnerable to disease (Duncan & Schaller, 2009). Studies have also documented links between perceived vulnerability to disease and overperceptions of unusual morphological features. For example, individuals higher in disease concerns set a lower threshold for categorizing someone as obese, and situationally priming disease threat leads to over-remembering seeing obese targets (Miller & Maner, 2012).

Members of other cultural groups may also be implicitly associated with disease threat. Human immune systems are attuned to local disease threats. Contact with unfamiliar outgroups might have historically increased the risk of contracting dangerous pathogens unfamiliar to locally-adapted immune systems (Diamond, 1999). An error management perspective predicts that the benefits of exaggerated avoidance of outgroup members (e.g. xenophobia) may have historically outweighed its costs (Kurzban & Leary, 2001). Indeed, individuals who tend to be particularly worried about infectious disease tend to hold more negative attitudes towards unfamiliar ethnic groups (Faulkner, Schaller, Park & Duncan, 2004), and making the threat of disease temporarily salient increases opposition towards policies allowing immigration of

unfamiliar outgroups (Faulkner et al., 2004). Ethnocentric and xenophobic attitudes are also higher for women during the first trimester of pregnancy, when the immune system is temporarily compromised (Naverette, Fessler, & Eng, 2007). Even cross-culturally, individuals in countries with higher levels of infectious disease are more likely to report that they would not want “people of a different race” as neighbors (Schaller & Murray, 2010).

Perceptions of Potentially Threatening People. Infectious disease is not the only threat posed by others, particularly for individuals who are physically vulnerable. A recent series of studies manipulated or measured vulnerability to harm and showed that vulnerable individuals overestimated the formidability of potentially threatening individuals. One study found that when people were told that a man was holding a gun, they perceived that person to be taller and more muscular than when they were told he was holding a drill, handsaw, or caulking gun (Fessler, Holbrook, & Snyder, 2012). Similarly, men who were temporarily physically incapacitated (either by being bound to a chair or by standing on a balance board) estimated an image of an angry man to be significantly taller and more muscular than did men who were not incapacitated (Fessler & Holbrook, 2013a). The presence of weapons also appears to influence dispositional judgments: Men who were pictured holding potentially harmful tools in non-violent situations (such as gardening shears) were judged to be more anger prone than when pictured holding innocuous tools (such as a watering can; Holbrook et al., 2014). In another study, parents with dependent children perceived a potentially threatening criminal to be more physically formidable than did non-parents (Fessler, Holbrook, Pollack, & Hahn-Holbrook, 2013).

Other situational variables can make potentially threatening individuals appear less physically formidable: Men who were in the presence of companions judged a solitary foe as

smaller and less muscular than men who made these judgments alone (Fessler & Holbrook, 2013b).

Biases in Interpersonal Perception

Sexual Overperception. Courtship communications are often ambiguous. Does a smile convey mere friendliness, or does it mean more? For men, error management logic predicts a bias toward overestimating a potential mate's sexual interest. This is because, all else equal, the reproductive costs of underestimating a woman's sexual interest and failing to pursue her – thereby missing out on an opportunity to reproduce – were likely to have been greater than the costs of pursuing a disinterested woman (Haselton & Buss, 2000). Men who were more successful in mating with greater numbers of women would have out-reproduced other men, passing along this possible overperception bias to their descendants.

For women, a different logic applies. Because of women's necessarily heavy investment in each child produced, and necessarily long interval between births, finding high quality partners – not more numerous partners – probably had a greater impact on women's reproductive success (Buss, 1994; Symons, 1979; Trivers, 1972). Therefore, error management logic predicts that men, but not women, possess the sexual overperception bias. Many sources of evidence support the sexual overperception hypothesis (see Haselton & Galperin, 2013, Table 11.1 for a review). For instance, in the earliest demonstration of the phenomenon, male and female strangers engaged in a get-to-know-you conversation in the lab and were viewed by a second pair of male and female strangers through one-way glass (Abbey, 1982). Both the male participant in the conversation (target) and the male observer rated the female target as more flirtatious and sexually interested than the female observer and female target. In this study and similar later studies, the difference between male and female ratings of women's sexual interest

was present when men's ratings were compared to the target woman's self-ratings and when compared to ratings made by third-party women assessing the interaction (Abbey, 1982; Haselton & Buss, 2000).

Similar results are found in surveys of men's and women's misperception experiences. In one study, for example, women reported more instances within the last year in which men overestimated their sexual interest than in which men underestimated it, suggesting that these men overperceived women's sexual interest in naturalistic situations outside of the lab (Haselton, 2003). Men in the same study reported roughly equal numbers of overperception and underperception errors on the part of women, providing no evidence of a bias in women. In related studies of opposite-sex friendships, men estimated their female friends' sexual interest to be greater than those women reported it to be (Koenig, Kirkpatrick, & Ketelaar, 2007; Bleske-Rechek et al., 2012). A recent speed-dating study similarly found that men estimated greater sexual interest in their female partners than their female partners reported (Perilloux, Easton, & Buss, 2012). This study also assessed variation in this apparent bias and found that men who were higher in short-term mating orientation and who were higher in self-rated attractiveness had a larger bias. Further, men's apparent bias was greater when they interacted with relatively more attractive women. In some birds, insects, and mammals (Alcock, 1993, chapter 13; Domjan, Huber-McDonald, & Holloway, 1992), males sometimes attempt to copulate with objects that only vaguely resemble females of their species, such as beer bottles or crude female models, suggesting similar behavioral biases in other species.

Commitment Skepticism. The reverse asymmetry might have applied to ancestral women as they estimated men's intentions to commit to long-term relationships (Haselton & Buss, 2000). Inferring long-term commitment interest in a man in whom it was absent could have

resulted in abandonment after the woman had already conceived a child, a high cost error potentially associated with reduced offspring survival (e.g., Hurtado & Hill, 1992).

Underestimating a man's commitment could also be costly, including delays in reproduction, but these costs might have been lower on average than costs associated with desertion (Haselton & Buss, 2000). Women might therefore possess a bias toward underestimating men's interest in commitment. Consistent with this idea, several studies have shown that women rate men's commitment given various courtship behaviors, such as giving gifts and verbal affirmations of love, lower than men rate it (Haselton & Buss, 2000). In contrast, women and men tend to agree on the level of commitment indicated by women's inaction of the same behaviors (Haselton & Buss, 2000). A recent study similar to earlier research documented apparent commitment skepticism in women who were prior to the age of menopause, but not women past the age of menopause, possibly because women past the age of reproduction would not have faced the same reproductive costs of overestimating men's commitment (Cyrus, Schwarz, & Hassabraubck, 2011).

Further evidence for commitment underperception in women was found in a study of face-to-face interactions between previously unacquainted male-female dyads (Henningsen & Henningsen, 2010). Dyads engaged in a five-minute conversation and afterwards filled out questionnaires about their own and their partner's perceived level of interest in a committed long-term relationship. Consistent with the commitment underperception hypothesis, women estimated lower levels of commitment interest than men reported for themselves. In contrast, men's estimates of women's commitment were not significantly different from women's reports of their commitment interest, providing no evidence of bias in men's judgments of women.

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 ill-tempered (Quillian & Pager, 2001) than are members of their own group. This bias might have been adaptive for reasons that extend beyond those related to the threat of disease transmission, presented above. 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 about aggression, particularly for inferences about out-group members who are not part of an individual's regular social circle. Schaller and colleagues proposed that cues indicating 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). Indeed, people who completed a rating task in a darkened laboratory endorsed racial and ethnic stereotypes connoting violence more so than people completing the task in a brightly lit room (Schaller et al, 2003). Darkness had no effect on other negative stereotypes of out-group others (such as laziness or ignorance) (Schaller et al, 2003).

Social Exchange Bias. Behavioral economists have puzzled over the fact that people cooperate in economic games with economic incentive structures favoring defection (Sally, 1995; Caporael et al. 1989; Camerer & Thaler 1995; Henrich et al. 2001). 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 economic tasks.

One possibility is 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, Mifune & Kanazawa, 2007). According to this logic, the costs of falsely believing one can defect without negative social consequences are often higher than cooperating when one 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 higher than in the more individualist United States samples (Yamagishi, et al., 1999). Also consistent with the social exchange bias hypothesis, when people 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 led to think this (Yamagishi et al, 2007; see also Savitsky, Epley, & Gilovich, 2001, and Williams, Case & Govan, 2003, for related predictions). Similar predictions were tested using evolutionary modeling and showed that one-shot cooperation can evolve due to the asymmetric costs of sometimes mistaking a repeated interaction for one-shot relative to mistaking one-shot for repeated interaction (Delton, Krasnow, Cosmides, & Tooby, 2011).

Note that this bias can be conceptualized as some combination of error management, as in the social exchange bias account, and an artifact of modern living, since in an ancestral environment the probability of re-encountering individuals would have been high and social reputation effects very potent. Thus, people may be predisposed to expect negative consequences of non-prosocial behavior even when, objectively, such consequences are unlikely to follow. The

bias towards prosociality has been the subject of competing explanations that take quite different explanatory stances (Price, Cosmides and Tooby, 2002; Gintis et al., 2003; Henrich & Boyd 2001; Bowles & Gintis, 2002), although these explanations might not be mutually exclusive.

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 (2006). Here we briefly discuss the representative example of the ‘positive illusions’.

Positive Illusions and Unrealistic Optimism. These are a well-known cluster of findings in judgement 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. One explanation is that 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 et al. 2003).

An alternative explanation involves error management. Nettle (2004) outlines such an explanation, building upon the interpretation of the positive illusions given by Taylor and Brown (1989). In evaluating a possible behavior, there are two possible errors. One may judge that the behavior is worthwhile when in fact it achieves nothing to promote fitness, or judge that a behavior is not worthwhile when in fact it enhances fitness to do it. The former error (a false positive) leads to behaviors that are actually 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, especially in competitive contexts. Thus, evolution can be expected to produce mechanisms biased towards 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). Recent neuroscientific research suggests that these biases have deep cognitive roots: individuals tend to encode undesirable information in a distorted manner, which leads to the relative enhancement of desirable information (Sharot, Korn, & Dolan, 2011; Sharot, Riccardi, Raio, & Phelps, 2007). Note that the error management account does not predict blanket optimism, but optimism where fitness gains were potentially high relative to the cost of passivity.

A similar argument can be made for a different type of positive illusion: overconfidence. Although overconfidence can sometimes lead to costly decisions and behaviors, its motivational benefits—in the form of increased ambition and persistence—might outweigh these costs. Evolutionary models are consistent with the notion that biased representations of personal success probabilities can be favored under certain circumstances (Johnson & Fowler, 2011; but see Marshall et al., 2013 and Johnson & Fowler, 2013).

Other evolutionary models suggest that states that are overconfident in warlike behaviors are more likely to be successful than accurate or underconfident states (Johnson, Weidmann, & Cederman, 2011).

Conclusions

Research on cognitive and social bias has been dominated by the failure and bleak implications of heuristics (see Kruger & Funder, 2004). A *Newsweek* magazine account of the

heuristics and biases literature 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 et al. 1999, p. 27). In reflecting back 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, 2004; Haselton et al., 2009).

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. Natural selection is similarly responsible for the intricacy of the human mind. 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 has now turned. There has been a shift toward explanations for bias invoking adaptive function, 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 in fact 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 many different 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, auditory looming and navigation biases, and overestimation of physical formidability of threatening targets).

Recent investigations have also begun to document conditions that moderate certain error management biases. Further investigations of the sexual overperception bias, for example, have shown that, along with sex of the perceiver predicting overperception (the classic EMT finding), that the perceiver's levels of interest predicted even greater apparent sexual overperception (e.g., Koenig et al., 2007). Similarly, in studies of overperception and commitment skepticism in face-to-face interactions, whereas sexual overperception only emerged among men who were sexually interested in the women they were interacting with, commitment skepticism was reduced or eliminated among women who were sexually interested in the men they were interacting with (Henningsen & Henningsen, 2010). Investigations such as these provide an increasingly nuanced understanding of biases predicted by error management theory.

Many questions remain. Some scholars have noted that a *cognitive bias* is not actually necessary to manage error costs—a wholly accurate *cognitive evaluation* coupled with *behavioral bias* could be equally effective or superior to a cognitive bias (McKay & Dennett,

2009; McKay & Efferson, 2010). Consider sexual overperception. A man does not need to have a *biased belief* that a woman is sexually interested in order to approach her. He might think to himself, “My chances are low, but why not try?”

This is a plausible alternative design for managing error costs. Although EMT was originally advanced to explain cognitive biases, the core logic of the theory is neutral in predicting whether a bias must be built into belief or occur further along in the decision chain, leading more directly to biased actions. The question of whether solutions to error management problems are sometimes rooted in biased *belief* is an open question that can only be answered on a case-by-case basis with empirical research (Haselton & Buss, 2009). However, as sexual misperception biases, perceptual auditory looming biases, navigation biases, and many others demonstrate, there is abundant evidence that people’s beliefs are indeed biased. Therefore, the argument that, in theory, error management adaptations need not involve biased beliefs does not render true cognitive biases nonexistent or impossible. The state of the evidence indicates otherwise. The fascinating puzzle that remains is an explanation for why humans often seem to have biased beliefs when a behavioral bias might suffice. One possibility is that the functional thinking that has guided error management theory will need to be more fully integrated with an understanding of the proximate mechanisms that give rise to biases (Marshall et al., 2013). Such an integration could reveal that the easiest or most effective way for an evolved brain to deliver behavioral biases is via cognitive biases (Haselton & Buss, 2009).

In sum, the notion that human judgment is fundamentally flawed appears to have been flawed itself. 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 might also prove to be more functional than one would think at first.

Some genuine cognitive 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*, 1621-30.
- 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.
- 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*. Basic Books.
- Bleske-Rechek, A. Somers, E. Micke, C. Erickson, L. Matteson, L., Stocco, C., Schumacher, B. & Ritchie, L. (2012). Benefit or burden? Attraction in cross-sex friendship. *Journal of Social and Personal Relationships*, *29*, 569–596.
- Bowles, S. & Gintis, H. (2002) Homo reciprocans. *Nature* *415*, 125-8.
- 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 der Kragt, A.J. (1989). Selfishness examined. *Behavioral and Brain Sciences* 12, 683-739.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, 31, 187-276.
- Cosmides, L., Barrett, H. C., & Tooby, J. (2010). Adaptive specializations, social exchange, and the evolution of human intelligence. *Proceedings of the National Academy of Sciences*, 107, 9007-9014.
- 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.
- Cyrus, K., Schwarz, S., & Hassebrauck, M. (2011). Systematic cognitive biases in courtship context: Women's commitment–skepticism as a life-history strategy? *Evolution and Human Behavior*, 32, 13-20.
- Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, 108, 13335-13340.
- Diamond, J. M. (1999). *Guns, Germs and Steel*. New York: W.W. Norton & Co.
- Domjan, M., Huber-McDonald, M., & Holloway, K. S. (1992). Conditioning copulatory behavior to an artificial object: Efficacy of stimulus fading. *Animal Learning & Behavior*. 20, 350-362.

- Duncan, L. A., & Schaller, M., (2009). Prejudicial attitudes toward older adults may be exaggerated when people feel vulnerable to infectious disease: Evidence and implications. *Analysis of Social Issues and Public Policy*, 9, 97-115.
- 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.
- Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes and Intergroup Behavior*, 7, 333–353.
- Fessler, D. M., & Holbrook, C. (2013a). Bound to Lose: Physical Incapacitation Increases the Conceptualized Size of an Antagonist in Men. *PloS one*, 8, e71306.
- Fessler, D. M., & Holbrook, C. (2013b). Friends shrink foes: The presence of comrades decreases the envisioned physical formidability of an opponent. *Psychological Science*, 24, 797-802.
- Fessler, D. M., Holbrook, C., Pollack, J. S., & Hahn-Holbrook, J. (2014). Stranger danger: Parenthood increases the envisioned bodily formidability of menacing men. *Evolution and Human Behavior*, 35, 109-117.
- Fessler, D. M., Holbrook, C., & Snyder, J. K. (2012). Weapons make the man (larger): Formidability is represented as size and strength in humans. *PloS one*, 7, e32751.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123-129.
- Fodor, J. A. (2001). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT press.

- Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science, 17*, 1068-1074.
- Garcia, J., Hankins, W. G., & Rusiniak, K. W. (1976). Flavor aversion studies. *Science, 192*, 265-266.
- Gerbert, B., Sumser, J., & Maguire, B. T. (1991). The impact of who you know and where you live on opinions about AIDS and health care. *Social Science and Medicine, 32*, 677-681.
- Gigerenzer, G. (1997). Ecological intelligence: An adaptation for frequencies. *Psychologische Beitrage, 39*, 107-129.
- Gigerenzer, G., Todd, P. M. & the ABC Research Group (1999). Simple heuristics that make us smart. New York: Oxford University Press.
- Gigerenzer, G., & Wegwarth, O. (2013). Five-year survival rates can mislead. *BMJ*, doi: 10.1136/bmj.f548.
- Gintis, H., Bowles, S., Boyd, R. & Fehr, E. (2003). Explaining altruistic behavior in humans. *Evolution and Human Behavior 24*, 153-72.
- Goldsmith, T.H. (1990). Optimization, constraint and history in the evolution of eyes. *Quarterly Review of Biology, 65*, 281-322.
- Goodwin, S. A., Gubin, A. Fiske, S. T., & Yzerbyt, V. T. (2000). Power can bias subordinates by default and by design. *Group Processes & Intergroup Relations, 3*, 227-256.
- Green, D. M. & Swets, J. A. (1966). *Signal detection and psychophysics*. New York: Wiley.
- Greene, K., & Banerjee, S. C. (2006). Disease-related stigma: Comparing predictors of AIDS and cancer stigma. *Journal of Homosexuality, 50*, 185-209.

- 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., Bryant, G. A., Wilke, A., Frederick, D. A., Galperin, A., Frankenhuis, W. E., & Moore, T. (2009). Adaptive rationality: An evolutionary perspective on cognitive bias. *Social Cognition, 27*, 733-763.
- 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, NY: Cambridge.
- Haselton, M. G., & Buss, D. M. (2009). Error management theory and the evolution of misbeliefs. *Behavioral and Brain Sciences, 32*, 522-523.
- Haselton, M. G., & Galperin, A. (2013). Error management in relationships. *Handbook of Close Relationships, 234-254*.
- Haselton, M. G., & Nettle, D. (2006). The paranoid optimist: An integrative evolutionary model of cognitive biases. *Personality and social psychology Review, 10*(1), 47-66.
- Hawkes, K. (2003). Grandmothers and the evolution of human longevity. *American Journal of Human Biology 15*: 380-400.
- Henningsen, D. D., & Henningsen, M. L. M. (2010). Testing error management theory: Exploring the commitment skepticism bias and the sexual overperception bias. *Human Communication Research, 36*, 618-634.

- 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., others. (2001). Cooperation, reciprocity and punishment in 15 small-scale societies. *American Economic Review* 91, 73-78.
- Herek, G. M. (1999). AIDS and stigma. *American Behavioral Scientist*, 42, 1106–1116.
- 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.
- Holbrook, C., Galperin, A., Fessler, D. M., Johnson, K. L., Bryant, G. A., & Haselton, M. G. (2014). If looks could kill: Anger attributions are intensified by affordances for doing harm. *Emotion*, 14, 455-461.
- Houston, V., & Bull, R. (1994). Do people avoid sitting next to someone who is facially disfigured? *European Journal of Social Psychology*, 24, 279-284.
- Hurtado, A. M., Hill, K. R. (1992). Paternal effect on offspring survivorship among Ache and Hiwi hunter-gatherers. In B. S. Hewlett et al. (Eds.), *Father-child relations: Cultural and biosocial contexts*. (pp. 31-55). New York: Aldine De Gruyter.
- Inhorn, M. C., & Brown, P. J. (1990). The anthropology of infectious disease. *Annual Review of Anthropology*, 19, 89-117.

Jackson, R. E., & Cormack, L. K. (2007). Evolved navigation theory and the descent illusion.

Perception & Psychophysics, *69*, 353–362.

Johnson-Laird, P., Legrenzi, P. & Legrenzi, M. (1972). Reasoning and a sense of reality. *British*

Journal of Psychology *63*, 495-500.

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: error management, cognitive constraints, and adaptive decision-making biases.

Trends in Ecology & Evolution, *28*, 474-481.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*, 317-320.

Johnson, D. D. P. & Fowler, J. H. (2013). Complexity and simplicity in the evolution of decision-making biases. *Trends in Ecology and Evolution*, *28*, 446–447.

Johnson, D. D. P., Weidmann, N. B. & Cederman, L.-E. (2011). Fortune favours the bold: An agent-based model reveals adaptive advantages of overconfidence in war. *PLoS ONE*, *6*, e20851.

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

Keltner, D. Gruenfeld, D. H, & Anderson, C. (2003). Power, approach, and inhibition.

Psychological Review, *110*, 265-284

Kenrick, D. T., Neuberg, S. L., Giskevicius, V., Becker, D. V., & Schaller, M. (2010). Goal-Driven Cognition and Functional Behavior: The Fundamental-Motives Framework.

Current Directions in Psychological Science, *19*, 63-67.

- Koenig, B. L., Kirkpatrick, L. A., & Ketelaar, T. (2007). Misperception of sexual and romantic interests in opposite-sex friendships: Four hypotheses. *Personal Relationships, 14*, 411-429.
- Krueger, J., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27*, 313-328.
- Kurzban, R., Leary, M. R. (2001). Evolutionary origins of stigmatization: The functions of social exclusion. *Psychological Bulletin, 123*, 187-208.
- Lieberman, D. L., Tybur, J. M., & Latner, J. D. (2012). Disgust sensitivity, obesity stigma, and gender: Contamination psychology predicts weight bias for women, not men. *Obesity, 20*, 1803-1814.
- Marshall, J.A.R., Trimmer, P. C., Houston, A. I., & McNamara, J. M., (2013). On evolutionary explanations of cognitive biases. *Trends in Ecology & Evolution, 28*, 469–473.
- McKay, R. T., & Dennett, D. C. (2009). The evolution of misbelief. *Behavioral and Brain Sciences, 32*, 493-510.
- McKay, R., & Efferson, C. (2010). The subtleties of error management. *Evolution and Human Behavior, 31*, 309-319.
- Mellers, B., Hertwig, R. & Kahneman, D. (2001). Do frequency representations eliminate conjunction effects? An exercise in adversarial collaboration. *Psychological Science, 12*, 269-275.
- Miller, S. L., & Maner, J. K. (2012). Overperceiving disease cues: The basic cognition of the behavioral immune system. *Journal of personality and social psychology, 102*, 1198-1213.

- Navarrete, C. D., Fessler, D. M. T., & Eng, S. J. (2007). Elevated ethnocentrism in the first trimester of pregnancy. *Evolution and Human Behavior, 28*, 60–65.
- Nettle, D. (2004). Adaptive illusions: Optimism, control and human rationality. In D. Evans & P. Cruse (Eds.), *Emotion, Evolution and Rationality*. Oxford: Oxford University Press.
- Nettle, D. and M. Bateson (2012). Evolutionary origins of mood and its disorders. *Current Biology, 22*, 712-721.
- Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. *Ecological Psychology, 13*, 87-110.
- Neuhoff, J. G., Hamilton, G. R., Gittleson, A. L., & Mejia, A. (2014). Babies in traffic: Infant vocalizations and listener sex modulate auditory motion perception. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 775-783.
- Neuhoff, J. G., Long, K. L., & Worthington, R. C. (2012). Strength and physical fitness predict the perception of looming sounds. *Evolution and Human Behavior, 33*, 318-322.
- Park, J. H., Faulkner, J., & Schaller, M. (2003). Evolved disease-avoidance processes and contemporary anti-social behavior: Prejudicial attitudes and avoidance of people with physical disabilities. *Journal of Nonverbal Behavior, 27*, 65–87.
- Park, J. H., Schaller, M., & Crandall, C. S. (2007). Pathogen-avoidance mechanisms and the stigmatization of obese people. *Evolution and Human Behavior, 28*, 410–414.
- Perilloux, C., Easton, J. A., & Buss, D. M. (2012). The misperception of sexual interest. *Psychological Science, 23*, 146-151.
- Pinker, S. (2005). So how does the mind work?. *Mind & Language, 20*, 1-24.
- Price, M., Cosmides, L. & Tooby, J. (2002). Punitive sentiment as an anti-free rider psychological adaptation. *Evolution and Human Behavior 23*, 203-31.

- Proffitt, D. R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review*, 2, 409-428.
- Quillian, L. & Pager, D. (2001). Black neighbors, higher crime? The role of racial stereotypes in evaluations of neighborhood crime. *American Journal of Sociology* 107, 717-67.
- 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.
- 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 & Social Psychology*, 81, 44-56.
- Schaller, M., & Murray, D. R. (2010). Infectious diseases and the evolution of cross-cultural differences. In M. Schaller, A. Norenzayan, S. J. Heine, T. Yamagishi, & T. Kameda (Eds.), *Evolution, culture, and the human mind* (pp. 243-256). New York: Psychology Press.
- 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.

- Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature neuroscience, 14*, 1475-1479.
- Sharot, T., Riccardi, A. M., Raio, C. M., & Phelps, E. A. (2007). Neural mechanisms mediating optimism bias. *Nature, 450*, 102-105.
- Stefanucci, J. K., & Proffitt, D. R. (2009). The roles of altitude and fear in the perception of height. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 424-438.
- Stefanucci, J. K., Proffitt, D. R., Clore, G. L., & Parekh, N. (2008). Skating down a steeper slope: Fear influences the perception of geographical slant. *Perception, 37*, 321-323.
- 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.
- Tooby, J., & Cosmides, L. (1990). On the universality of human nature and the uniqueness of the individual: The role of genetics and adaptation. *Journal of Personality, 58*, 17-67.
- 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, IL: Aldine Publishing Company.
- 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.
- Voland, E. & Beise, J. (2002). Opposite effects of maternal and paternal grandmothers on infant survival in historical Krummhorn. *Behavioral Ecology and Sociobiology, 52*, 435-443.

- Wason, P. C. (1983). Realism and rationality in the selection task. In J. Evans (Ed.) *Thinking and Reasoning: Psychological approaches*. London: Routledge & 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, 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.
- Witt, J. K., & Sugovic, M. (2013). Spiders appear to move faster than non-threatening objects regardless of one's ability to block them. *Acta Psychologica*, 143, 284-291.
- 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., Mifune, N., & Kanazawa, S. (2007). The social exchange heuristic: Managing errors in social exchange. *Rationality and Society*, 19, 259-291.
- Zebrowitz, L. A., & Montepare, J. (2006). The ecological approach to person perception: Evolutionary roots and contemporary offshoots. In M. Schaller, J. A. Simpson, & D. T. Kenrick (Eds.), *Evolution and social psychology* (pp. 81–113). New York: Psychology Press.

Table 1: Evolutionary Taxonomy of Cognitive Biases

| <i>Type of Bias</i> | <i>Examples</i> |
|--|---|
| <i>Heuristic</i> : Bias results from evolutionary or information processing constraints; mechanisms work well in most circumstances, but are prone to break down in systematic ways. | <ol style="list-style-type: none"> 1. Use of stereotypes 2. Fundamental attribution “error” 3. One-reason decision strategies |
| <i>Artifact</i> : 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. | <ol style="list-style-type: none"> 1. Some instances of base-rate neglect in statistical prediction 2. Some instances of the confirmation bias |
| <i>Error Management Bias</i> : Selection favored bias toward the less costly error; although error rates are increased, net costs are reduced. | <ol style="list-style-type: none"> 1. Auditory looming 2. Xenophobia 3. Sexual overperception by men 4. Commitment underperception by women |