PART I

FUNDAMENTAL INFLUENCES ON SOCIAL JUDGMENTS

2

Biases in Social Judgment

Design Flaws or Design Features?

Martie G. Haselton and David M. Buss

CONTENTS

INTRODUCTION

Humans appear to fail miserably when it comes to rational decision making. They ignore base rates when estimating probabilities, commit the *sunk cost* fallacy, are biased toward confirming their theories, are naively optimistic, take undue credit for lucky accomplishments, and fail to recognize their self-inflicted failures. Moreover, they overestimate the number of others who share their beliefs, demonstrate the *hindsight bias*, have a poor conception of chance, perceive illusory relationships between noncontingent events, and have an exaggerated sense of control. Failures at rationality do not end there. Humans use external appearances as an erroneous gauge of internal character, falsely believe that their own desirable qualities are unique, can be induced to remember events that never occurred, and systematically misperceive the intentions of the opposite sex (for reviews, see Fiske & Taylor, 1991; Kahneman, Slovic, & Tversky, 1982; and Nisbett and Ross, 1980; for cross-sex misperceptions of intentions, see Haselton & Buss, 2000). These documented phenomena have led to the

We thank Paul Andrews and Julie Smurda for helpful comments on portions of this chapter.
Please address correspondence to Martie Haselton: haselton@ucla.edu

widespread conclusion that our cognitive machinery contains deep defects in design.

This conclusion has not gone unchallenged (also see Brewer, this volume, and Funder, this volume). Some suggest that certain documented irrationalities are artifacts of inappropriate experimental design (e.g., Cosmides & Tooby, 1996; Gigerenzer, 1996). Others argue that the normative standards against which human performance is compared are inappropriate (Cosmides & Tooby, 1994; Fox, 1992; Pinker, 1997). Recently, we have articulated Error Management Theory (EMT), which proposes that some biases in human information processing should not be viewed as flaws at all (Haselton & Buss, 2000). To understand why demonstrating *bias* does not logically entail *flaw* requires knowledge of the general causal process responsible for fashioning human cognitive mechanisms and the specific adaptive problems humans were designed to solve.

THE HEURISTICS AND BIASES APPROACH

The study of cognitive biases in social psychology can be traced to the creative and influential work of Kahneman and Tversky (1972, 1973; Tversky & Kahneman, 1971, 1973, 1974). In their studies, Kahneman and Tversky documented surprisingly flagrant violations of basic rules of probability. The famous "Linda problem" (Tversky & Kahneman, 1983) is illustrative. Subjects in the Linda studies were provided with a short 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 also 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 conjunct proposition cannot be more likely than either of its constituent elements, between 80% and 90% of subjects tend to select (b) as the more probable option, committing what Tversky and Kaheman (1983) called the *conjunction fallacy*.

Kahneman, Tversky, and others following in the *heuristic-and-biases* tradition documented many such violations, including neglect of base rates, misconceptions of chance, illusory correlation, and anchoring bias (Tversky & Kahneman, 1974; see Shafir & LeBoeuf, 2002, for a recent review). Theoretically, these purported irrationalities have been explained as a necessary consequence of the mind's limited computational power and time. As Tversky and Kahneman explain, "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 EVOLUTIONARY FOUNDATIONS OF SOCIAL JUDGMENTS AND DECISIONS

There are no known scientific alternatives to evolutionary processes as causally responsible for shaping organic mechanisms. There are no compelling arguments that humans have been exempt from this causal process. Nor is there reason to believe that human cognitive mechanisms have been outside the purview of evolution by natural selection.

The premise that human cognitive mechanisms, at some fundamental level of description, are products of the evolutionary process, however, does not by itself provide the information required for knowing precisely what those mechanisms are. There is wide disagreement, even among evolutionary theorists, about the nature of the products of the evolutionary process, especially when it comes to humans (e.g., Alexander, 1987; Buss, 1995; Tooby & Cosmides, 1992).

The primary disagreement centers on the relative domain specificity versus domain generality of the evolved mechanisms (Kurzban & Haselton, in press). At one end of the conceptual spectrum, some have argued for versions of domain-general rationality – that human consciousness has the power to "figure out" what is in the individual's best fitness interest. At the other end are those who argue that evolution by selection has produced a large and complex array of specific psychological mechanisms, each designed to solve a particular adaptive problem (e.g., Buss, 1991; Symons, 1987; Tooby & Cosmides, 1992). According to this line of theorizing, highly domain-general mechanisms are unlikely to have evolved, in part because they lead to *combinatorial explosion* – the rapid multiplication of potential alternative ways of cleaving the perceptual environment and of selecting courses of action (Tooby & Cosmides, 1992).

Humans have evolved to solve specific adaptive problems – avoiding predators, keeping warm, eating food, choosing mates – in real time. What constitutes a successful solution in one domain differs from successful solutions in other domains. Criteria for successful food selection (e.g., rich in calories and nutrients, lacking in toxins), for example, differ radically from criteria for successful mate selection (e.g., healthy, not already mated). One all-purpose mechanism is generally inefficient, and sometimes massively maladaptive, for solving adaptive problems that differ widely in their criteria for successful solution. Because there are only small islands of successful adaptive solutions, selection tends to favor specialized mechanisms that prevent drowning in the vast sea of maladaptive ones (Tooby & Cosmides, 1992).

This theoretical orientation has important implications for conceptualizing human information processing machinery. It suggests that the appropriate criterion against which human judgment is evaluated should not be the abstract, content-free principles of formal logic (Cosmides, 1989). Rather, human rationality should be evaluated against a different criterion – whether the information processing mechanism succeeds, on average, in solving the relevant adaptive problem.[1] Because what constitutes a successful solution will differ across domains, no single standard can in principle be appropriate for evaluating human judgment. Indeed, the chapters in this volume provide much evidence of functionally distinct processes surrounding social ostracism (Williams, Case, & Govan, this volume), attachment (Shaver & Mikulincer), and intergroup stereotyping effects (Galinski, Martorana, & Ku, this volume), to name a few. Perhaps most importantly, this theoretical perspective suggests that the most successful adaptive solutions, for some adaptive problems, are those that are systematically biased.

In the balance of this chapter, we will show how this principle applies to considerations of the appropriateness of research designs and the selection of normative standards in the heuristics-and-biases approach. In the section on normative standards, we highlight EMT (Haselton & Buss, 2000), a new perspective on the evolution of social biases.

THE FORMAT AND CONTENT OF ADAPTIVE PROBLEMS

Ecologically Relevant Problem Formats

Data from nonhuman organisms with neurological systems considerably simpler than those of humans adhere closely to the same rules of probability humans are proposed to violate (Cosmides & Tooby, 1996). Foraging behavior in bumblebees, for example, adheres to some rules of probability (Real, 1991), and similarly sophisticated statistical logic has been identified in birds (Real & Caraco, 1986). Moreover, evidence from the study of language (Pinker & Bloom, 1992), visual perception (Shepard, 1992), and many other areas within human psychology suggests that the human mind does indeed possess computationally sophisticated and complex information processing mechanisms. If neurologically and computationally modest brains can embody a calculus of probability, why not human brains too? If other computational systems within the human mind are functionally

[1] Other constraints suggested by the evolutionary approach may also apply to this evaluation. There may be considerable differences between the current and past conditions under which the mechanism evolved. Moreover, in addition to limits on information processing time and capacity (Tversky & Kahneman, 1974), there are constraints imposed by the operation of multiple information processing systems, some of which may have competing demands (e.g., Lieberman, this volume). Thus, *evolutionary rationality* should be based on whether the solution works, somehow, in both solving a problem and coordinating with other mechanisms, and given the alternatives available at the time, and given the material constraints, and eventually leading (in the past) to successful replication.

complex, why should we not expect reasonably good performance involving assessments of probabilities?[2]

One possibility is that the mismatch between human performance and Bayesian expectations is an artifact of inappropriate experimental design. On evolutionary-ecological grounds, Gigerenzer (e.g., 1991, 1997) proposed that tasks intended to assess whether human reasoning embodies laws of probability should present information in a frequency format rather than in probabilities, as is typical in heuristics-and-biases tasks.

He argues that if information was represented in a stable format over human evolutionary history, mental algorithms designed to use the information can be expected to operate properly only when presented with information in that format, even if an alternative format is logically equivalent. Although numerical information can be represented equally well in binary and base-10 form, for example, a pocket calculator will produce logical output only when the input is in base 10. Statistically formatted probabilities are an evolutionarily novel format for computing event likelihood. Some kinds of natural frequencies, on the other hand, are easily observed and have been recurrently available over evolutionary history. For example, one can easily note the number of occasions on which one has met John and he has behaved aggressively versus those occasions on which he did not. According to this logic, if we wish to see whether humans can use Bayesian logic (e.g., inferring the likelihood of events given certain cues), we should present information in frequency form.

As predicted by this account, frequency formats reliably improve performance in tasks like the Linda problem (for sample problems, see Table 2.1). Whereas the probability format produces violations of the conjunction rule in 50% to 90% of subjects (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983), frequency formats decrease the rate of error to between zero and 25% (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983). Cosmides and Tooby (1996) documented a similar effect by rewording the medical diagnosis problem (Casscells, Shoenberger, & Graboys, 1978) to shift it toward the frequency format.

The literature preceding these challenges is vast, with a large constituency, so naturally, these startling results are controversial (Gigerenzer, 1996; Kahneman & Tversky, 1996; Mellers, Hertwig, & Kahneman, 2001). Nevertheless, the frequency effect appears to be reliable (see the preceding discussion), and it cannot be attributed to a simple clarification of the terms involved in the original problems (Cosmides & Tooby, 1996, Exp. 6) or to the addition of the extensional cues (Hertwig & Gigerenzer, 1999, Exp. 4)

[2] See Cosmides and Tooby (1996) for further elaboration of arguments suggesting that a calculus of probability should be expected in human judgment if the relevant adaptive problems recurred over human evolutionary history.

TABLE 2.1. *Comparison of Frequency and Probability Formats in the Linda Problem*

Representation	Conjunction Violations
Probability Format	
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 also participated in antinuclear demonstrations.	82–88% (across two studies)
Rank the following hypotheses according to their probability: Linda is a bank teller. Linda is a bank teller and active in the feminist movement.	
Frequency Representation	
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. [. . .]	20%
Imagine 200 women who fit the description of Linda. Please estimate the frequency of the following events. How many of the 200 women are bank tellers? —— out of 200 How many of the 200 women are bank tellers and active in the feminist movement? —— out of 200	
Frequency Variant	
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. [. . .]	16%
Imagine women who fit the description of Linda. Please estimate the frequency of the following events. How many of the women are bank tellers? —— out of —— How many of the women are bank tellers and active in the feminist movement? —— out of ——	

Source: Adapted from Hertwig and Gigerenzer (1999).

implicated when performance improved in earlier studies (Tversky and Kahneman, 1983).

On the other hand, even with greatly reduced rates of error in the frequency format, some evidence suggests a lingering bias toward conjunction errors over other sorts of errors for subjects who fail to solve the conjunction problem correctly (Tversky & Kahneman, 1983). Thus, the frequency studies do not rule out the possibility that people use systematically fallible heuristics in solving some problems. To argue this point is, however, to miss the critical insight to be gained from these studies. The key point is that the human mind can use a calculus of probability in forming judgments, but to observe this, one must present problems in evolutionarily valid forms.

Evolutionarily Relevant Problem Content

A similar conclusion emerges from evolutionary-ecological research on the Wason selection task (Cosmides, 1989; Cosmides & Tooby, 1992; Fiddick, Cosmides, & Tooby, 2000). Past studies with the task suggested that people are unable to use proper falsification logic (Wason, 1983). Revised versions of the studies in which falsification logic was required to detect cheating in social contracts (Cosmides, 1989) or avoid dangerous hazards (Pereya, 2000) caused performance to increase from rates lower than 25% correct (Wason, 1983) to over 75% correct (Cosmides, 1989). The researchers argued that performance increased so dramatically because past studies used highly abstract rules that failed to tap into evolutionarily relevant problem domains. The revised studies did so and thereby activated evolved problem-solving machinery that embodies proper falsification logic (Cosmides, 1989; Cosmides & Tooby, 1992; Pereya, 2000; Fiddick et al., 2000).

The conclusion to be drawn from these studies is *not* that humans actually 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, human performance can be shown to improve greatly.

In summary, shifting the format or content of problems of social judgment toward greater adaptive relevance can sometimes greatly improve performance. In the context of this volume, these changes might be thought of as shifts from *high* processing to *deep* processing (Brewer, this volume). High processing may be susceptible to limitations imposed by processing ability (Stanovich & West, 2000), effort, and motivation (Kruglanski et al., this volume), whereas deep processing may be relatively more effortless and unintentional, and it may often be superior because it is that for which selection has created specific adaptive design.

WHAT COUNTS AS GOOD JUDGMENT?

Adaptive versus Truthful Inferences

An evolutionary perspective raises questions about what should count as a good judgment. It suggests that the human mind is designed to reason adaptively, not truthfully or even necessarily rationally (Cosmides & Tooby, 1994). The criterion for selection is the net fitness benefit of a design relative to others that happen to be visible to selection at the time. Sometimes this might produce reasonably truthful representations of reality, whereas at other times it might not.

As Pinker notes, "conflicts of interest are inherent to the human condition, and we are apt to want *our version* of the truth, rather than the truth itself to prevail" (1997, p. 305; emphasis in the original). Thus, it might be for good adaptive reasons that we tend to overestimate our contributions to joint tasks (Ross & Sicoly, 1979), have positively biased assessments of ourselves (Brown, 1986), and believe that our strongly positive qualities are unique but that our negative ones are widely shared by others (Marks, 1984).

Biased Trade-offs

Biases can also emerge as a consequence of trade-offs. All adaptations have costs as well as benefits. Cost–benefit trade-offs can produce reasoning strategies prone to err in systematic ways (Arkes, 1991; Tversky & Kahnman, 1974). Less often recognized is the proposal that trade-offs in the relative costs of errors can produce biases (Haselton & Buss, 2000). It is this potential insight to which we now turn.

ERROR MANAGEMENT THEORY

Errors and Bias in Social Signal Detection Problems

Understanding and predicting the behavior of others is a formidable social task. Human behavior is determined by multiple factors, people sometimes mislead others for their own strategic purposes, and many social problems require inferences about concealed events that have already occurred or future events that might occur. It is unavoidable, therefore, that social judgments will be susceptible to error. Given the necessary existence of errors, how should these systems best be designed?

At one level of abstraction, we can think of two general types of errors in judgment: false positives and false negatives. A decision maker cannot simultaneously minimize both errors because decreasing the likelihood of one error necessarily increases the likelihood of the other (Green & Swets, 1966). When the consequences of these two types of errors differ in their relative costliness, the optimal system will be biased toward committing the less costly error (also see Cosmides & Tooby, 1996; Friedrich, 1993; Nesse & Williams, 1998; Schlager, 1995; Searcy & Brenowitz, 1988; Tomarken, Mineka, & Cook, 1989; Wiley, 1994).

Consider a human-made device as an example. Smoke alarms are designed to be biased toward false-positive errors, because the costs of missing an actual fire are so much more severe than the relatively trivial costs of putting up with false alarms. Similarly, the systems of inference in scientific decision making are biased, but in the reverse direction, because many scientists regard false positives (type I errors) as more costly than false negatives (type II errors).

In the smoke alarm example, the designer's (and buyer's) intuitive evaluation of what constitutes cost and benefit guides the design of the system. (Or perhaps those smoke alarm makers who designed systems that produced an equal number of false positives and false negatives went out of business.) The evolutionary process, however, provides a more formal calculus by which competing decision-making mechanisms are selected – the criterion of relative fitness. If one type of error is more beneficial and less costly than the other type of error, in the currency of fitness, then selection will favor mechanisms that produce it over those that produce less beneficial and more costly errors, even if the end result is a larger absolute number of errors. One interesting conclusion from this line of reasoning is that a *bias*, in the sense of a systematic deviation from a system that produces the fewest overall errors, should properly be viewed as an *adaptive bias*.

In sum, when the following conditions are met, EMT predicts that human inference mechanisms will be adaptively biased: (1) when decision making poses a significant signal detection problem (i.e., when there is uncertainty); (2) when the solution to the decision-making problem had recurrent effects on fitness over evolutionary history; and (3) when the aggregate costs or benefits of each of the two possible errors or correct inferences were asymmetrical in their fitness consequences over evolutionary history.

In our research, we have used EMT to predict several social biases. To start with, we hypothesized that there might exist biases in men's and women's interpretation of courtship signals. Because of their ambiguity and susceptibility to attempts at deception, these signals are prone to errors in interpretation. Moreover, with their close tie to mating and reproduction, courtship inferences are a likely target for adaptive design. Based on EMT, we advanced two hypotheses: the sexual overperception hypothesis and the commitment skepticism hypothesis.

Sexual Overperception by Men

We proposed that men possess evolved inferential adaptations designed to minimize the cost of missed sexual opportunities by overinferring women's sexual intent (Haselton & Buss, 2000). One primary factor limiting men's reproductive success over evolutionary history was their ability to gain sexual access to fertile women (Symons, 1979). Ancestral men who tended to infer falsely a prospective mate's sexual intent paid the relatively low costs of failed sexual pursuit – perhaps only some lost time and wasted courtship effort. In contrast, men who tended to infer falsely that a woman lacked sexual interest paid the costs of losing a reproductive opportunity. Given that sexual access to fertile women was one of the most important limiting resources for men, and a missed sexual opportunity could have meant some likelihood of missing a direct opportunity to reproduce, it is reasonable to hypothesize that the social inference errors that resulted in missed sexual opportunities were more costly than the social awkwardness associated with overestimating sexual interest.

Many studies, using a variety of different methods, are consistent with and have supported predictions based on this hypothesis. In laboratory studies, when male partners in previously unacquainted male–female dyads are asked to infer their partner's sexual interest, they consistently rate it as higher than the female partner's report suggests and higher than the ratings provided by female third-party viewers of the interaction (Abbey, 1982; Saal, Johnson, & Weber, 1989). A similar effect occurs in studies using photographic stimuli (Abbey & Melby, 1986), videos (Johnson, Stockdale, & Saal, 1991), short vignettes (Abbey & Harnish, 1995), ratings of courtship signals (Haselton & Buss, 2000), and in surveys of naturally occurring misperception events (Haselton, in press). Importantly, evidence of sexual overperception does not appear in women (Haselton & Buss, 2000; Haselton, in press). (Figure 2.1 presents a representative effect.)[3]

Commitment Underperception by Women

We proposed an opposing inferential bias in women (Haselton & Buss, 2000). For women, the costs of falsely inferring a prospective mate's commitment when little or none exists were probably greater than the costs of failing to infer commitment that does exist. An ancestral woman who consented to sex with a man she intended to be her long-term mate, but who subsequently failed to invest in her and her offspring, could have suffered the costs of an unwanted or untimely pregnancy, raising a child without an investing mate, and possibly a reduction in her mate value (Buss, 1994). These were substantial costs given the lowered survival of the child (Hurtado & Hill, 1992) and impairment of future reproductive potential. An ancestral woman who erred by underestimating a man's commitment, in contrast, may have merely evoked more numerous and more frequent displays of commitment by the man who truly was committed. If, as we

[3] For nonhuman animal examples consistent with this proposal, see Alcock (1993) on *sexually indiscriminant behavior*.

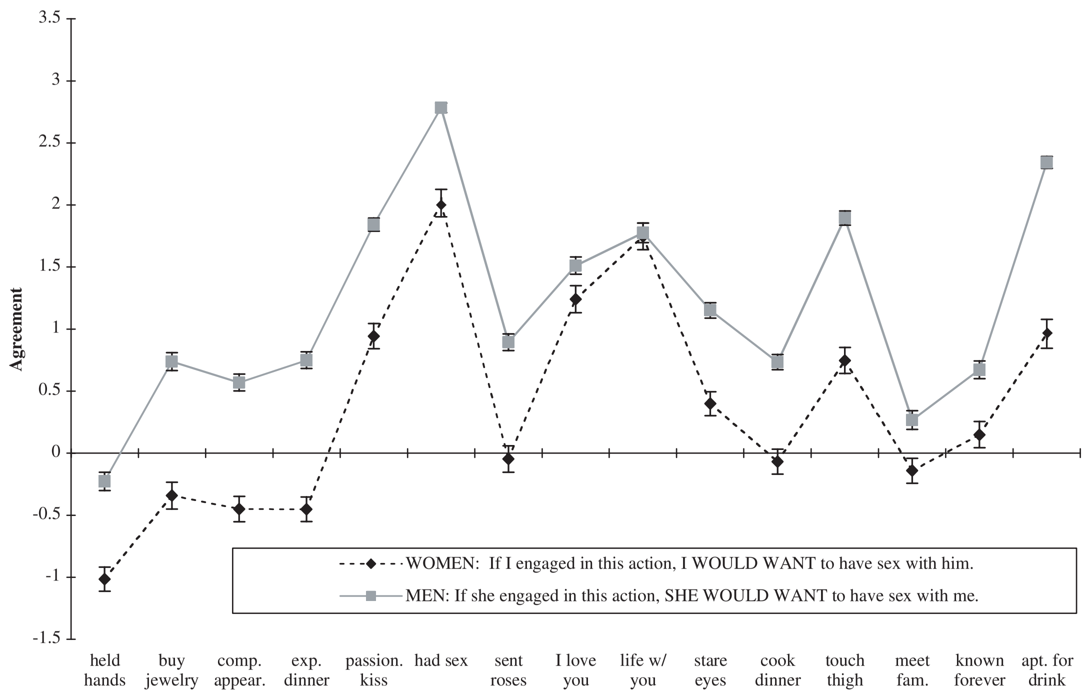

Figure 2.1. Comparison of women's and men's assessments of the degree to which courtship signals indicate sexual interest in women. Women ($n = 168$) provided ratings of their own likely sexual intent, and men ($n = 121$) provided ratings of a female dating partner's likely sexual intent, as determined by each dating behavior. All differences are significant at the $\alpha = .05$ level except "saying she [I] would like to spend her [my] life with me [him]." Men's ratings of women's sexual interest also exceed women's ratings of other women's sexual interest (see Haselton & Buss, 2000, Study 1 and Study 2, Part 2); thus, the difference does not appear to be a simple self–other difference attributable to self-serving biases. In the same studies, women's ratings of men's sexual interest did not deviate systematically from men's ratings (Haselton & Buss, 2000), suggesting that the sexual overperception effect is limited to male perceivers. Error bars represent standard errors. (Adapted from Haselton & Buss, 2000.)

propose, false-positive errors in inferring commitment were more costly than false-negative errors for ancestral women, selection should have favored commitment-inference mechanisms in women biased toward underinferring men's commitment.

In our studies, we have found evidence of the predicted effect. In the context of courtship, women infer that potential indicators of men's desire for a committed relationship (e.g., verbal displays of commitment and resource investment) indicate less commitment than men report that they intend them to indicate (Haselton & Buss, 2000; see Figure 2.2). The same result appears when comparing women's and men's ratings of a third-party man's commitment signals (Haselton & Buss, 2000). Importantly, evidence of commitment bias does not appear in men's assessments of women's signals (Haselton & Buss, 2000).

Overperception of the Risk of Aggression and Homicide

We have also hypothesized that recurrent cost asymmetries have produced a bias toward overinferring aggressive intentions in others (Haselton & Buss, 2000). Duntley and Buss (1998) tested this hypothesis in the context of homicide. They asked men and women to estimate their own likelihood of committing homicide under certain conditions (e.g., if given a million dollars to do so). They also asked them how likely they were to be victims of homicide under the same conditions (if someone else was given a million dollars to kill them). There was a large and significant difference between these assessments; people estimated that their own likelihood of killing was far lower (10%) than their likelihood of being killed (80%). Although the possibility of a self-serving response bias cannot be ruled out by these data alone, these results are consistent with the error management prediction that we may overestimate the aggressive intentions of others to avoid the high costs of injury or death (Haselton & Buss, 2000).

Other Error Management Effects

Snake Fear. It is likely that the costs of different errors in inferring danger about snakes were asymmetrical over human evolutionary history, with failures to flee clearly being the more costly error (Tomarken et al., 1989). Many studies support the expectations of this hypothesis. Snake fears are easily acquired relative to other fears, they are notoriously difficult to extinguish, and even when successful, discrimination training quickly tends to revert back to a generalized snake fear (Mineka, 1992). Subjects in studies using a covariation paradigm tend to overestimate the association of negative outcomes such as shock and ancestrally dangerous stimuli (snakes or spiders), but do not overestimate the association of shock and innocuous stimuli (mushrooms or flowers; Tomarken et al., 1989), nor do

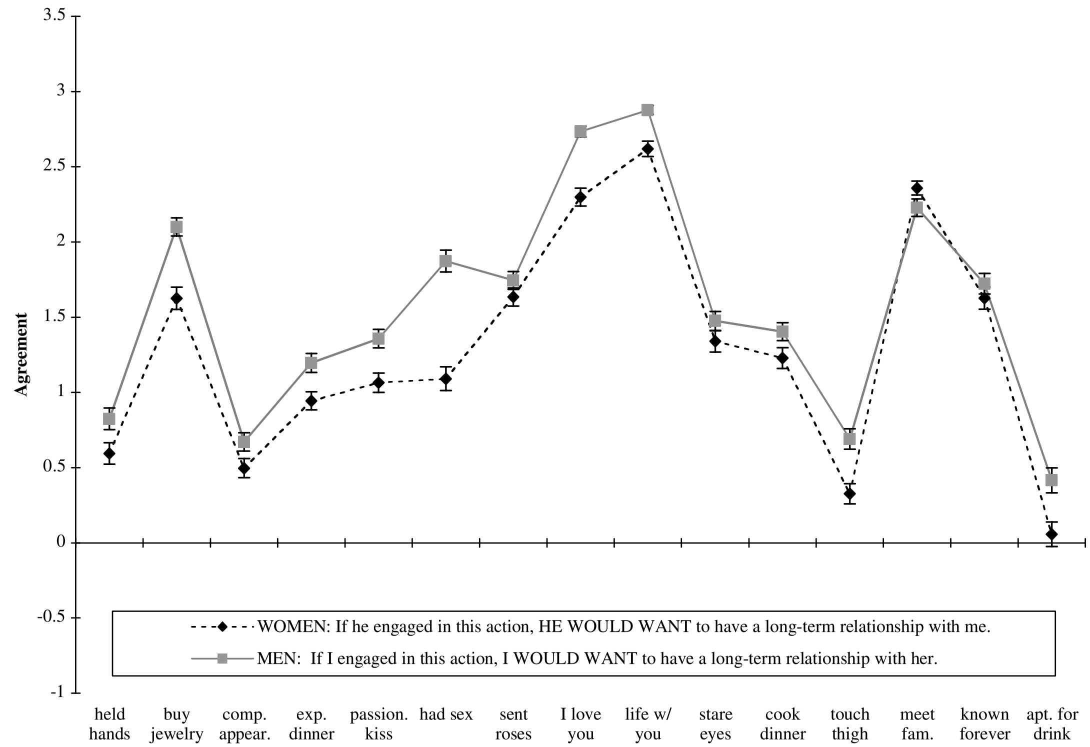

Figure 2.2. Comparison of men's and women's assessments of the degree to which courtship signals indicate commitment intent in men. Men ($n = 121$) provided ratings of their own likely commitment, and women ($n = 168$) provided ratings of a male dating partner's likely commitment, as determined by each dating behavior. All differences are significant at the $\alpha = .05$ level except "stared deeply into her [eyes], "cooked her [me] a gourmet dinner," and told her [me] I [he] would like to meet her [my] family." Women's ratings of men's commitment also exceed men's ratings of other men's commitment (see Haselton & Buss, 2000, Study 1 and Study 2, Part 2); thus, the difference does not appear to be a simple self–other difference attributable to self-serving biases. In the same studies, men's ratings of women's commitment did not deviate systematically from women's ratings (Haselton & Buss, 2000), suggesting that the commitment skepticism effect is limited to female perceivers. Error bars represent standard errors. (Adapted from Haselton & Buss, 2000.)

they overestimate the association of shock and presently dangerous stimuli with strong semantic associations with shock, such as electrical outlets (Tomarken et al., 1989).

Food Preferences and Disgust. Humans are finicky about what comes into contact with their food and are not willing to take many chances when deciding whether to consume something. Even though very young children are proposed to lack the cognitive prerequisites for understanding invisible contaminants, most 3- and 4-year-olds refuse to drink juice into which a sterilized cockroach has been dipped (Siegal & Share, 1990). Adults will not drink from a brand new urine-collection cup, even if the researcher takes it from a box of brand-new cups from the factory (Rozin & Fallon, 1987). People are even reluctant to sweeten a beverage with sugar taken from a bottle that says "not poison" or "not sodium cyanide," which suggests that the mere mention of contamination is sufficient to render a food inedible (Rozin, Markwith, & Ross, 1990). In ancestral conditions, in times when food was relatively abundant, the costs of failing to infer correctly that something was contaminated or poisonous were undoubtedly higher than the costs of inferring that a safe food was dangerous. Our present-day conditions typically represent such abundance and, as error management logic suggests, people appear to be biased against consuming foods that might be contaminated, even if the possibility is remote.

Application of EMT to Known Social Biases

Might EMT also help to explain some of the classic effects social psychologists have been interested in, such as the fundamental attribution error, illusory correlation, and the like? We suggest two particularly intriguing possibilities.

Paranoid Social Cognition. Paranoid social cognition occurs when people feel a misplaced or exaggerated sense of distrust and suspicion that often leads to inferences that others are negatively evaluating or conspiring against them (Kramer, 1998). A particular form of social paranoia is the *sinister attribution error* (Kramer, 1994). For example, if a colleague were to pass by in the hallway without saying hello, her coworker could assume that (1) she was simply busy and did not see him or (2) she disliked him or thought he was unimportant. The sinister attribution error occurs if construals of potentially negative events are biased toward personalistic interpretations like the latter.

From an error management perspective, this bias could make sense if it helped people to avoid missing highly costly situations in which they actually were the potential victims of vengeful, disdainful, or conspiratorial others. Of course, the vigilance that such a disposition would promote might distract too much attention from other important activities and prove too costly to maintain if it characterized all social inferences. As one might expect from this additional line of logic, the bias does not occur in all contexts. It is exhibited more by individuals who are under intense scrutiny, new to social organizations (e.g., first-year graduate students), or lower in status than others (e.g., assistant professors) (Kramer, 1994, 1998). These may be analogous to the ancestral conditions in which individuals may have incurred the greatest level of social risk, such as when immigrating into a new village or tribe with apprehensive inhabitants or when joining a new coalition.

The Fundamental Attribution Error. The fundamental attribution error (FAE) occurs when people assume that the behaviors of others and their mental contents (attitudes, dispositions) correspond to a greater degree than is logically warranted (Andrews, 2001; Gilbert & Malone, 1995; Ross, 1977). There are many potential causes of the effect (e.g., Andrews, 2001; Gilbert & Malone, 1995). One possibility is that the nature of the two causal categories under consideration (correspondent vs. noncorrespondent causes) yields different levels of prediction for the perceiver.

Correspondent causes, such as the fact that an aggressor's behavior was caused by his aggressive disposition or intense dislike of a person, are relatively limited in number. Noncorrespondent causes for the aggressor's behavior, such as the fact that the aggressor was provided money by an external source or was otherwise coerced to aggress, are potentially infinite in number (Andrews, 2001). Moreover, correspondent causes of behavior are easy to locate (they reside within the individual), whereas noncorrespondent causes will often be linked to situational factors that can vary, in principle, in an infinite number of ways (Funder, 2001). All else equal, it might simply pay off more for a perceiver to assume that the cause of behavior is the target's disposition or attitude (and that the person will behave aggressively again) than to attribute the behavior to any given element of a situation, even if 95% of the variance in behavior is due to situational determinants.

From an error management perspective, the FAE may occur if underattribution of behaviors to corresponding mental states leads to more costly errors than overattribution (Andrews, 2001). This proposal is particularly compelling when the behavior in question could have highly negative consequences for a perceiver who failed to protect her- or himself from its perpetrator. Prime examples are aggressive, jealous, uncooperative, or politically distasteful behaviors. As Andrews (2001) suggests, it may be for precisely this reason that past histories of criminal convictions are inadmissible as evidence in court proceedings – they bias jurors too much. "In the hunter-gatherer groups that we evolved in, an individual with a history of uncooperative behavior was likely to pose a stable threat to the interests of other group members. If such a person was under suspicion for committing another offense, other group members might have had an interest in biasing their assessments toward guilt because the costs of erroneous inference of guilt were low and the costs of an erroneous inference of innocence were potentially high" (Andrews, 2001, p. 25).

CONCLUSIONS

The social world is replete with uncertainty. Was that sound in the dark a random accident or was it produced by a lurking intruder? Does this smile indicate sexual interest or mere friendliness? Who can I trust? Who will betray me? Amid the chaos of social cues, humans must render judgments, generate inferences, make decisions, and solve tangible problems in real time. Social psychology has taken on the important scientific task of identifying, describing, and explaining the information processing machinery responsible for producing social solutions to these problems.

Evaluating the performance of cognitive mechanisms requires the specification of standards against which performance is compared. Much work in the field has used content-free principles of formal logic as the crucial criteria and has found human performance wanting. This work, in the aggregate, has illuminated something important – that human judgment and inference can be shown to deviate, sometimes dramatically, from the optimal solutions that would be produced by formal logical systems (but see Funder, this volume, for a discussion of the *error paradigm*).

An evolutionary perspective suggests an alternative set of evaluative criteria – those anchored in successful solutions to specific and delimited adaptive problems. Evolution by selection does not favor information processing machinery according to whether it follows formal logic or even whether it maximizes truthful inferences (Cosmides & Tooby, 1994; Haselton & Buss, 2000). Nor does it favor content-general rules of information processing over content-specific ones (Symons, 1992). It favors mechanisms that minimize overall costs and maximize overall benefits, in the currency of evolutionary fitness, relative to the analogous costs and benefits generated by alternative designs present at the time.

There are several implications of this perspective that we believe have important consequences for the interpretation of biases. First, deviations from normative standards can be produced when problems are not presented in evolutionarily valid formats such as natural frequencies. Toasters should not be found deficient for failing to brown food presented in the shape of potatoes. Analogously, human cognition should not be found deficient for failing at tasks presented in forms that have no evolutionary precedent and for which the mind is not designed.

Deviations can also be produced when subjects are asked to solve abstract problems with evolutionarily novel content far removed from the adaptive problems human mechanisms were designed to solve. Thus, whether human reasoning is good or bad, and whether it adheres to formal standards or not, often cannot be known from studies that have simply demonstrated bias. The format of the input conditions and the problem content also must be scrutinized. Moreover, as Funder (1987, this volume) notes, bias in laboratory studies may or may not correspond to real-world mistakes, particularly when a subject's performance is deemed erroneous unless it lands squarely at the predicted point of optimal performance – a performance standard that should be missed due to measurement error alone.

Second, demonstrations of bias should not necessarily be explained by invoking notions of information processing constraints, cognitive shortcuts, or corner-cutting heuristics. Error management theory points to an alternative explanation – that cognitive biases are sometimes favored by natural selection because of the net benefit they produce relative to alternative designs. These biases, far from being mere by-products of constraints, are adaptive features of good design.

In summary, many have concluded that documented irrationalities suggest widespread flaws in human psychological design. These biases have been called *fallacies*, (Kahneman & Tversky, 1972), *illusions*, and *sins* (Piatelli-Palmarini, 1994) and have been described as "ludicrous" and "indefensible" (Tversky & Kahneman, 1971). An evolutionary perspective suggests alternative criteria for evaluating the information processing machinery that drives social inference.

EMT, as a recent approach, has been explored in only a few domains – the sexual overperception bias displayed by men, the commitment skepticism bias displayed by women, and the aggressive intent bias exhibited by both sexes. It is too early to assess whether EMT can help to explain a wider array of biases that psychologists have discovered, from the hindsight bias to the FAE. But it does suggest a principled way in which these biases can be reexamined in a new conceptual light. And it provides a theoretical tool with which new biases can be discovered.

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.

Abbey, A., & Harnish, R. J. (1995). Perception of sexual intent: The role of gender, alcohol consumption, and rape supportive attitudes. *Sex Roles*, *32*, 297–313.

Abbey, A., & Melby, C. (1986). The effects of nonverbal cues on gender differences in perceptions of sexual intent. *Sex Roles*, *15*, 283–289.

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

Alexander, R. D. (1987). *The biology of moral systems*. Hawthorne, NY: Aldine.

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.

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

Brown, J. D. (1986). Evaluations of self and others: Self enhancement biases in social judgments. *Social Cognition*, *4*, 353–376.

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

Buss, D. M. (1995). Evolutionary psychology: A new paradigm for psychological science. *Psychological Inquiry*, *6*, 1–30.

Casscells, W., Schoenberger, A., & Grayboys, T. B. (1978). Interpretation by physicians of clinical laboratory results. *New England Journal of Medicine*, *299*, 999–1001.

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

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York: Oxford University Press.

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.

Duntley, J., & Buss, D. M. (1998, July). *Evolved anti-homicide modules*. Paper presented at the annual meeting of the Human Behavior and Evolution Society Conference, Davis, CA.

Fiddick, L., Cosmides, L., & Tooby, J. (2000). No interpretation without representation: The role of domain specific representations and inferences in the Wason selection task. *Cognition*, *77*, 1–79.

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

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition (2nd ed.)*. New York: McGraw-Hill.

Friedrich, J. (1993). Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena. *Psychological Review*, *100*, 298–319.

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

Funder, D. G. (2001). The really, really fundamental attribution error. *Psychological Inquiry*, *12*, 21–23.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. In W. Stroebe & M Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 129–162). New York: Wiley.

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

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

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, *117*, 21–38.

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

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.

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

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, C. B., Stockdale, M. S., & Saal, F. E. (1991). Persistence of men's misperceptions of friendly cues across a variety of interpersonal encounters. *Psychology of Women Quarterly*, *15*, 463–475.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

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.

Kramer, R. M. (1994). The sinister attribution error: Paranoid cognition and collective distrust in organizations. *Motivation and Emotion*, *18*, 199–230.

Kramer, R. M. (1998). Paranoid cognition in social systems: Thinking and acting in the shadow of doubt. *Personality and Social Psychology Review*, *2*, 251–275.

Kurzban, R. O., & Haselton, M. G. (in press). Making hay out of straw: Real and imagined controversies in evolutionary psychology. In J. Barkow (Ed.), *Missing the revolution: Darwinism for social scientists*. New York: Oxford University Press.

Marks, G. (1984). Thinking one's abilities are unique and one's opinions are common. *Personality and Social Psychology Bulletin*, *10*, 203–208.

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

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

Nesse, R. M., & Williams, G. C. (1998). Evolution and the origins of disease. *Scientific American*, *11*, 86–93.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Pereya, L. (2000, June). *Function variation of the hazard management algorithm*. Paper presented at the Human Behavior and Evolution Society Conference, Amherst, MA.

Piatelli-Palmarini, M. (1994). *Inevitable illusions*. New York: Wiley.

Pinker, S. (1997). *How the mind works*. New York: W. W. Norton.

Pinker, S., & Bloom, P. (1992). Natural language and natural selection. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 451–493). New York: Oxford University Press.

Real, L. (1991). Animal choice and the evolution of cognitive architecture. *Science*, *253*, 980–986.

Real, L., & Caraco, T. (1986). Risk and foraging in stochastic environments: Theory and evidence. *Annual Review of Ecology and Systematics*, *17*, 371–390.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York: Academic Press.

Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*, 322–336.

Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review*, *94*, 23–41.

Rozin, P., Markwith, M., & Ross, B. (1990). The sympathetic magical law of similarity, nominal realism, and neglect of negatives in response to negative labels. *Psychological Science*, *1*, 383–384.

Saal, F. E., Johnson, C. B., & Weber, N. (1989). Friendly or sexy? It may depend on whom you ask. *Psychology of Women Quarterly*, *13*, 263–276.

Schlager, D. (1995). Evolutionary perspectives on paranoid disorder. *Delusional Disorders*, *18*, 263–279.

Searcy, W. A., & Brenowitz, E. A. (1988, March 10). Sex differences in species recognition of avian song. *Nature*, *332*, 152–154.

Shafir, E., & LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology*, *53*, 491–517.

Shepard, R. N. (1992). The three-dimensionality of color: An evolutionary accommodation to an enduring property of the natural world? In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 495–532). New York: Oxford University Press.

Siegal, M., & Share, D. L. (1990). Contamination sensitivity in young children. *Developmental Psychology*, *26*, 455–458.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, *23*, 645–726.

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

Symons, D. (1987). If we're all Darwinians, what's all the fuss about? In C. B. Crawford, M. F. Smith, & D. L. Krebs (Eds.), *Sociobiology and psychology: Ideas, issues, and applications* (pp. 121–146). Hillsdale, NJ: Erlbaum.

Symons, D. (1992). On the use and misuse of Darwinism. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 137–159) New York: Oxford University Press.

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

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

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*, 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

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.

Wason, P. C. (1983). Realism and rationality in the selection task. In J. Evans (Ed.), *Thinking and reasoning: Psychological approaches* (pp. 44–75). London: Routeledge & Kegan Paul.

Wiley, R. H. (1994). Errors, exaggeration, and deception in animal communication. In L. A. Real (Ed.), *Behavioral mechanisms in evolutionary ecology* (pp. 157–189). Chicago: University of Chicago Press.