The Evolution of Accuracy and Bias in Social Judgment

Martie G. Haselton

UCLA

David C. Funder

UC Riverside

Chapter in preparation for M. Schaller, J. A. Simpson, & D. T. Kenrick (eds.) *Evolution and Social Psychology*, Psychology Press.

Draft: December 1, 2004

The Evolution of Accuracy and Bias in Social Judgment

Humans are an intensely social species and therefore it is essential for our interpersonal judgments to be valid enough to help us to avoid enemies, form useful alliances and find suitable mates; flawed judgments can literally be fatal. An evolutionary perspective implies that humans ought to have developed sufficient skills at solving problems of interpersonal judgment, including gauging the personalities of others, to be useful for the basic tasks of survival and reproduction. Yet, the view to be derived from the large and influential bias-and-error literature of social psychology is decidedly different—the social mind seems riddled with fundamental design flaws. We will argue in this paper that flawed design is probably the least plausible explanation for the existence of so many errors. We present an evolutionarily-based taxonomy of known bias effects that distinguishes between biases that are trivial or even artifactual and lead virtually nowhere, and those that have interesting implications and deserve further study. Finally, we present an evolutionary perspective that suggests that the ubiquity, automaticity, and success of interpersonal judgment, among other considerations, presents the possibility of a universal *Personality Judgment Instinct*.

ADAPTATIONS FOR SOCIAL LIFE

Archeological evidence and behavioral patterns observed in extant hunter-gatherer groups indicate that the human species has been intensely social for a long time (e.g., Chagnon, 1983, Tooby & Devore, 1987). Human offspring have a remarkably extended period of juvenile dependency, which both requires and provides the skills for surviving in a complex social world (Hrdy, 1999). Humans evolved language and universal emotional expressions which serve the social purpose of discerning and influencing the thoughts of others (e.g., Darwin, 1872; Ekman, 1973; Pinker, 1994), and humans will infer social intentions on the basis of minimal cues, as Heider and Simmel (1944) demonstrated in their classic experiment involving chasing triangles and evading circles. Recent work has shown that children above age 4 and adults in disparate cultures (Germans and Amazonian Indians) can categorize intentions—chasing, fighting, following, playing, and courting (for adults)— from no more than the motion patterns of computerized v-shaped arrowheads (Barrett, Todd, Miller, & Blythe, in press).

Most notably, humans have a deeply-felt need for social inclusion. Deprivation of social contact produces anxiety, loneliness, and depression (Baumeister & Leary, 1995); indeed, as William James (1890) observed: “Solitary confinement is by many regarded as a mode of torture too cruel and unnatural for civilised countries to adopt." Participants in laboratory studies who are left out of a face-to-face triadic ball toss respond with depressed mood and decreased self esteem (Williams & Sommer, 1997). These effects can even be produced by a computerized version of the game in which participants use key presses to "toss" the ball back and forth to human-like figures on a screen (Williams, Cheung, & Choi, 2000), and persist when participants are told that the other players have been scripted or are mere computer programs (Zadro & Williams 2003, cited in Williams, Case, & Govan, 2003). Neuroscience evidence suggests that being ostracized activates the same brain regions involved in the sensation of physical pain (Eisenberger, Lieberman, & Williams, 2003). Rejection hurts, literally.

This acute social sensitivity makes sense in the light of the many problems of social adaptation that have long faced members of our species: the formation of cooperative alliances for hunting and protection (e.g., Tooby & Devore, 1987; **Tooby & Cosmides, 1988**), hierarchy negotiation (Kyl-Heku & Buss, 1996), mate choice (Buss, 2003; Miller, 2000; Symons, 1979; choice of allies and friends (Tooby & Cosmides, 1996), and social exchange (Cosmides 1989), to name a few. Given the importance of these problems, we should expect finely honed adaptations for forming social judgments and making social decisions that are, at the very least, good enough to promote survival and reproduction. We would certainly *not* expect thousands of years of social evolution to yield a psychological apparatus fundamentally prone to social misperception, judgmental flaws, and maladaptive interpersonal behavior. Yet, this is the picture one gets from a good deal of conventional research in social psychology.

IS THE SOCIAL MIND DEEPLY FLAWED?

A large part of social psychology – including some of its most famous and influential research programs – consists of a loosely-connected set of non-intuitive and curious effects, each of which demonstrates a context in which humans can be led to make incorrect judgments according to one or more standards of logic, statistics, or even morality (Krueger & Funder, in press). An especially famous error is the putative tendency for people to infer that dispositions (enduring aspects of personality) have stronger effects on the behavior of others than do situations, coined the *fundamental attribution error* (Ross, 1977; but see Funder, 1982). Humans also have been accused of false consensus, confirmation bias, overconfidence bias (as well as pessimistic bias), hindsight bias, and the sinister attribution error. And, experimenters have caught humans in the act of committing the planning fallacy, the external agency illusion, and the transparency illusion. These are just a few examples from a very long list (for reviews see Gilovich, Griffin, & Kahneman, 2002; Fiske & Taylor, 1991), a list that grows longer all the time. One recent example is the *dud-alternative effect* in which adding an implausible alternative (e.g., a "dud" in a horserace) increases the judged likelihood that a good alternative will win, when in fact the inclusion of more alternatives must reduce the probably of success for any given candidate (Windschitl & Chambers, 2004). An especially terrifying recent example is the *bias blind spot* (Pronin, Gilovich, & Ross, 2004), which is the bias to not know you're biased! The cumulative effect of this ever-growing list is a view of the human social mind as fraught with shortcomings, a view that is almost always detectable implicitly and often is expressed explicitly as well (Lopes, 1991).

The emphasis on bias and error is understandable to some degree. Bias effects tend to be counterintuitive (Lopes, 1991; Funder, 2003), funny—they make for good anecdotes and amusing classroom demonstrations (Crandall, 1984)—and to the degree they really do afflict cognition and associated life outcomes, they call out for study so that they can be fixed. However, the view of human judgment as dominated by error is both implausible and theoretically impoverished.

When a putative error of human judgment is discovered, there are three possible explanations. First, the error might not be an error at all. The experimental situation or instructions to subjects or the standards by which error has been defined might be misleading or incorrect, so that the putative error is better considered an experimental artifact. Second, the error might be one that, on balance, leads in realistic situations to adaptive decisions more often than not. The error might be produced by a usually-adaptive heuristic, or be the result of a tendency to favor less costly errors over more costly ones (see below). For example, to the extent that behavior really is predictable from stable traits and attitudes, the fundamental attribution error (to the extent it is itself not an artifact) will tend to produce correct decisions in realistic circumstances. Third, the error might reveal a flaw in psychological design such that the mind is fundamentally prone to get a broad class of decisions wrong. This explanation is the most frequently offered of the three but is, we submit, the least plausible one.

Furthermore, these loosely connected findings, despite their number, do not add up to a broad, coherent theory of human social thought and behavior. The reason is that demonstrations of error characteristically begin by assuming that human judgment to be perfect, and attain their news value from the conclusion that it is not. But this conclusion does not provide even the beginnings of an explanation of how judgments are ever made correctly; the initial assumption of perfection bypasses any possibility of a broader account. Instead, the long lists of errors powerfully convey the usually implicit, sometimes explicit and surely misleading message that good judgment is rarely achieved.

WHERE DO BIASES COME FROM?

Setting aside the possibility of fundamentally flawed mental design, we suggest that each of the many documented biases and errors may be (1) *artifacts* of inappropriate research strategies, and surprisingly many of the most famous ones may belong in this category, (2) may stem from *heuristics*, usually effective judgmental strategies that are subject to systematic breakdown, or (3) be the result of *error management*, a special case of a heuristic in which less costly errors are favored over more expensive ones (see Table 1; after Haselton, Nettle, and Andrews, in press). We consider each of these causes in turn.

Artifacts

Before beginning a serious analysis of the source of perceptual and judgmental bias, the first step is to set aside those that are little more than experimental artifacts. Researchers have found it easy to design artificial research settings in which individuals can be shown to err. But do such demonstrations reveal flaws in the design of the mind? Similarly, if the testing strategies researchers use to conduct research are more sensitive to error than to accuracy, people will appear error prone. But are they really?

Problem formats. Gigerenzer (1997) proposed that the human mind should be better at likelihood estimation when presented with information about discrete events as compared to numerical probabilities. Frequencies of events are what are observed in nature, he argued, whereas probabilities are invented, numerical abstractions that lack any direct connection to sensory input. Moreover, the computation of probabilities loses information about base rates (Cosmides & Tooby, 1996), so even if human sensory systems could take probabilities as input, frequencies may convey superior information.

In the famous Linda problem, subjects were asked to read a personality description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations." They were then asked to determine which of two options was more probable: (a) Linda is a bank teller, or (b) Linda is a bank teller and active in the feminist movement. Although the conjunction cannot be more likely that either of its constituents, between 80% and 90% of subjects select (b) as the more probable option, committing the "conjunction fallacy" (Tversky & Kaheman, 1983). However, simply changing the format of the problem from probabilities to frequencies (e.g., how many out of 200 women are bank tellers and how many out of 200 are bank tellers and feminists) dramatically improves performance (Fiedler, 1988; Hertwig & Gigerenzer, 1999; Tversky & Kahneman, 1983; Cosmides & Tooby, 1996, but see Mellers, Hertwig, & Kahneman, 2001; also see Hertwig & Gigerenzer, 1999 for related issues about violation of conversational norms in the classic problems). This insight has implications that go beyond identifying artifacts; Gigerenzer and his colleagues point out that in the relatively rare but important cases where exact probability judgments are important (e.g., in medical diagnosis), it is important to give decision-makers (e.g., doctors) relevant information in a form they can use (Hoffrage, Lindsey, Hertwig & Gigerenzer, 2000).

Problem content. Often researchers compare human performance to idealized rules of logic or specific statistical computations. Ironically, the very fact that it is feasible to present judgmental problems to humans using words, numbers and abstract concepts may be precisely why it so easy to demonstrate that people make mistakes in solving such problems. Imagine, by contrast, a study that tried to demonstrate fundamental irrationality, or even imprecise reasoning, in a dog. Would this be possible? We would guess not, which does not mean that canines are smarter than humans, at least not in most cases. Rather, the presentation of difficult abstract problems presented in complex verbal and numeric formats is something that can only inflicted on humans. You wouldn't do it to a dog. From an evolutionary perspective, the important problems of judgment are probably not very abstract. They involve estimations of others' specific motives and intentions (such as distinguishing friend from foe), predicting whether a patch of land will contain prey animals or predators, and, perhaps, detecting cheaters in social exchange.

As statistics teachers can attest, people do not find falsification logic either intuitively sensible or easy to employ. It comes as little surprise, then, that people are not especially good at testing the abstract conditional rule, if *p* then *q*. Wason (1983) showed that subjects correctly recognized that confirmatory evidence (the presence of *p*) is relevant to testing the rule, but they typically failed to test for falsifications of the rule (the absence of *q*). In the same line of research, however, a variety of content effects augmented performance on the task (e.g., Wason & Shapiro, 1971; Johnson-Laird, Legrenzi, & Legrenzi, 1972). Cosmides (1989) argued that many of these content effects reflect the operation of a cheater-detection algorithm. When the conditional rule involves social exchange (if you take the benefit [*p*] then you pay the cost [*q*]), people look not only for benefits taken (*p*) but also costs not paid (not *q*), increasing performance dramatically from 25% correct (Wason, 1983) to 75% correct (Cosmides, 1989). Similar effects are elicited by hazard-detection content, in which people are also induced to detect violations of a precaution rule about hazards (e.g., if you touch a contaminant [*p*] then you wash your hands [*q*]; Pereyra, 2000). Cosmides hypothesized that performance increases dramatically in these problems because the content elicited mechanisms for cheater detection or reasoning about hazards, both of which necessarily use falsification "logic" given the nature of the adaptive problem they are designed to solve (Cosmides, 1989; Pereya, 2000; Fiddick, Cosmides, & Tooby, 2000; see Cosmides & Tooby, 1992, for an extensive discussion, including a description of the many variants of the task devised to rule out confounds and alternative explanations). The conclusion to be taken from this work is not that humans actually do reason according to abstract rules of logic; in fact, results from the same line of work demonstrate cases in which adaptive responses systematically violate normative rules (see, e.g., the work on switched social contracts, Cosmides, 1989). Instead, the key message is that adaptive performance cannot be evaluated unless researchers present subjects with problems for which their minds are designed.

The Error Paradigm. The most basic reason to be skeptical of many of the putative demonstrations of error is that that the error paradigm, upon which most of these demonstrations are based, makes it extremely easy to detect error and almost impossible to detect accurate judgment. In the typical study, the normative response, whether it is derived from formal logic, abstract principles, or math, is a point prediction. For example, people have been asked to estimate other persons' attitudes, frequencies of behavioral compliance, probabilities of events, or even exact degrees of association between two variables. Only if subjects exactly attain the experimenter-defined correct estimate of the correct number will they be treated as accurate; needless to say, this almost never happens. The average estimate by real subjects will not exactly match the point-prediction; if enough subjects are in the sample, this deviation will be statistically significant, and a new bias will be born (Krueger & Funder, in press). The abundant evidence of error in the literature must be qualified by noting that the basic research strategy makes error easy to find.

Heuristics

As Herbert Simon famously observed, the best solutions to problems of judgment are often good enough—*satisficing*—rather than the best imaginable, because perfection may not be worth the extra cost (Simon, 1956). This observation inspired much of the most influential research on social judgment and decision making. Kahneman, Tversky and many others (see Gilovich, Griffin, & Kahneman, 2002, for a recent review) proposed that information processing time and cognitive capacity are limited and thus people use heuristics that trade-off accuracy for speed and efficiency. Surprisingly, however, many researchers forgot Simon's important message that such trade-offs still ought to yield decisions that are reasonably good. Instead, study after study was designed to show how the use of heuristics caused people to be, quite simply, wrong. For example, Tversky and Kahneman documented a variety of effects suggesting that humans do not use probability information properly.

For example, the "Linda problem" which we have already mentioned, led people to estimate Linda to be more likely to be a feminist bank teller than to be merely a bank teller. In studies in which subjects judged the likelihood of series of coin flips, subjects tended to say that HTHTTH was more likely than the sequence HHHTTT or HHHHTH, when in fact the former contains too many alternations and too few runs (Tversky & Kahneman, 1974). People seem to expect chance to be a self-correcting process (for a series of Hs to be corrected by a T), but of course each new flip is independent of the last, and in large samples correction has not occurred but rather repeated Hs or Ts have merely been diluted (Tversky & Kahneman, 1974).

Tversky and Kahnman attributed these effects and a variety of others to the use of mental short-cuts: "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). Indeed, when people are rushed in forming judgments, under cognitive load, or less motivated to be correct, their tendency to be biased (and presumably their use of heuristics) is more pronounced (see, e.g., Kahneman, 2003). Tversky and Kahneman (1974) offered the heuristic *representativeness* as an explanation for the conjunction fallacy and misconceptions of chance. When using the heuristic, people base their answers more on what is deemed representative of the category (feminist, random sequence) than on assessments of probabilities.

These investigators and others operating in the same paradigm have observed that they did not intend their findings to be taken as belittling the capacities of human judgment, and that heuristics such as representativeness might very well be part-and-parcel of good judgment under most circumstances. However, the consistent research strategy has been to show how these heuristics lead to errors, not how they ever, let alone typically, enhance accuracy or produce otherwise adaptive judgments. The widespread impression produced by this body of work, therefore, is of a massive number of studies demonstrating how heuristics produce errors, against few if any showing their adaptive possibilities. This impression has been encouraged by some of the rhetoric employed in research summaries (Lopes, 1991; Funder, 1992).

Only recently has this imbalance begun to be corrected. Gigerenzer, Todd, and colleagues (e.g., Gigerenzer, Todd, et al., 1999) developed a very different take on Simon's classic message. They observed that in addition to constraints imposed by time and cognitive capacity, the environment imposes quirky *informational* constraints (Gigerenzer, Czerlinski, & Martignon, 2002). People can only be expected to use information that is actually available to them in the current environment, or was so in the evolutionary environment in which decision-making strategies were forged. (As we discussed in the artifacts section, this perspective offers an alternative explanation for subjects' well-documented failure to use some forms of probability information correctly). People are also expected to possess strategies that exploit features of the informational environment leading to efficient—*fast-and-frugal*—decisions that are, as Simon would expect, valid enough to be useful (Gigerenzer, Todd, et al., 1999).

Gigerenzer and Goldstein (1996) showed that a family of simple decision-making rules that use only one datum can work as well or better than more complex algorithms (e.g., multiple regression) that use all possible information. An example is the recognition heuristic. When asked to make judgments about which of two alternatives will be higher on some criterion variable (such as who will succeed in a soccer game, or which city is larger) someone who uses the heuristic will choose the more familiar alternative. For example, when asked which city has a larger population, San Diego or San Antonio, German students tend to guess right: San Diego (Goldstein & Gigerenzer, 1999). American students tend to get this question wrong. This is the *less-is-more* effect—American students cannot use recognition since they have heard of both cities, so they rely on other cues that turn out to be invalid. The advantage reverses when German cities are compared; American students generally do better. Native residents know too much – both cities in their own country sound familiar to them – and therefore they cannot exploit the general principle that names of foreign cities are likely to seem more familiar if their population is larger.

The essential function of a heuristic is to guide someone who has little relevant information toward one or a few valid cues within a sea of possibilities. In this case, the cue is recognition. The heuristic would not be adaptive, and thus it would not persist as a feature of human cognition, if it did not produce useful decisions in most cases.

Error Management

The most common interpretation of biases involves trade-offs against constraining factors, such as time, cognitive resources, and the availability of information, as we have seen. But such an interpretation does not explain the particular *direction* of the bias exhibited. We suggest two possible solutions to this problem (also see Kenrick & Maner, in press; Krebs & Denton, 1997). The first is that the bias may actually serve, in most cases, to nudge inferences based on limited information in the direction of a valid or useful conclusion. For example, to the extent that human behavior really is affected by personality dispositions, an inference based on limited information that is biased in the direction of the fundamental attribution error is more likely to be correct than an inference not influenced by this bias. This process is analogous to the case in visual perception, where errors such as the Ponzo illusion or the Müller-Lyer illusion reveal mechanisms that cause limited 2-dimensional stimuli to be misjudged but allow correct judgment of size and distance in 3-dimensional contexts (Funder, 1987).

A second explanation for directional bias draws on the fact that judgments are not merely abstract outcomes; they are bases for action and therefore affect survival and reproductive success. As Kurzban and Aktipis (this volume) explain, the mind is not designed for logic or truth, per se. In some domains, such as representing certain aspects of the visual world, reasonable accuracy is adaptive, whereas in others, what is adaptively useful is might systematically misrepresent the truth.

Biases may, for example, be directed by trade-offs in error costs. Judgment mechanisms can make two general types of errors, false positives (false alarms) and false negatives (misses). For any given decision or judgment these two types of errors often differ in their costs. Sometimes a false alarm is highly costly. This is the case in scientific hypothesis testing, in which researchers have set the criterion for affirmation very high. In other cases a miss is more costly, as when people react to a threat too quickly but they are "better safe than sorry." Error Management Theory (Haselton & Buss, 2000, 2003; Haselton & Nettle, 2004) proposes that whenever the costs of errors in a given domain were consistently asymmetric over evolutionary history, judgment or decision-making adaptations should evolve to bias inferences toward the less costly error. Systems designed according to this engineering principle will tend to make more errors overall, but the errors will tend to be relatively cheap.

Haselton and Nettle (2004) argue that many apparent biases—from sensory perception to estimating the likelihood of events in the future—can be understood from this perspective. In perception, Neuhoff (2001) documented a*uditory looming* in estimations of the time-to-impact of approaching sounds. When people try to estimate the time of approaching sounds, they tend to underestimate the time of arrival, whereas when sounds move away, their estimates are unbiased (Neuhoff, 2001). Underestimation may have been favored by selection. When an object (such as a falling rock) is traveling toward you, it is better to anticipate its arrival too early than too late.

Haselton and Buss (2000) documented two error management effects in courtship communication. Abbey (1982) found that during brief cross-sex interactions men tended to rate women's sexual interest more highly than the women themselves did. Haselton and Buss (2000) proposed that this effect may reflect an evolved sexual overperception bias in men. They hypothesized that the fitness costs of underestimating a woman's sexual interest and thereby missing a sexual opportunity were greater on average than the costs of overestimating her interest and spending effort on fruitless courtship. Given women's selectiveness in mate choice (Trivers, 1972) and men's relatively greater willingness to engage in sex (e.g., Schmitt, et al., 2003), the same asymmetry does not hold for women's estimations of men's sexual intent. Three studies using diverse methods confirmed Abbey's original finding and showed, as predicted, that women do not show the same bias in interpreting men's sexual intent (Haselton & Buss, 2000; Haselton, 2003). In a recent set of experiments, Maner et al. (in press) found converging results. Men who were placed in a romantic frame of mind were particularly likely to see sexual arousal in women's facial expressions, especially when the women in the photographs were attractive. These results suggest that cues to increased reproductive benefits, which should further shift cost asymmetries to favor the false-positive bias, tend to yield corresponding increases in sexual overperception by men.

Haselton and Buss (2000) also predicted that women would be biased in interpreting men's courtship communications. They hypothesized that the fitness costs of overestimating a man's interest in forming a long-term relationship were greater ancestrally than the costs of underestimating it: the former could result in reproductive abandonment, whereas the latter would result in modest reproductive delays. As predicted, women appear to be commitment skeptical—relative to men, they tend to underestimate the degree of commitment conveyed by various dating actions (Haselton & Buss, 2000). Men show no such bias in interpreting women's commitment on the basis of the same cues (Haselton & Buss, 2000).

Many other biases may also be understood from the error management perspective. Defenses, such as allergy, cough, and anxiety should be somewhat over-responsive to threats (Nesse, 2001). Indeed, doctors can dampen these defenses with drugs and cause few untoward effects on their patients (Nesse, 2001). People may have a natural tendency to avoid diseased or injured persons to a greater extent than strictly necessary to avoid becoming ill themselves (Kurzban & Leary, 2001). There is evidence that this bias overextends to disabled individuals or individuals expressing phenotypic extremes (e.g., the obese) who pose no true threat (Park, Faulkner, & Schaller, 2003; Park, Schaller, & Crandall, 2004). Two sets of studies also indicate that cues linked with harm increase defensive biases. First, ambient darkness, a cue suggesting increased risk of hostility by others, increases subjects' stereotypes connoting violence in outgroup males, whereas other negative stereotypes do not change (Schaller, Park & Faulkner, 2003; Schaller, Park & Mueller, 2003). Second, subjects who are induced to feel fear in the laboratory see more anger in neutral facial expressions of outgroup males as compared with subjects induced to feel romantic arousal or those in the neutral emotion condition (Maner et al., in press).

"Positive illusions" (Taylor & Brown, 1988) may be understood as biases for promoting striving when the costs of expended efforts are lower than the costs of passivity, as trying and failing may not be very costly relative to failures to try at all (Nettle, 2004; also see Kurzban & Aktipis, this volume, and Taylor & Brown, 1988). For example, feeling optimistic and therefore increasing one's striving for uncertain fitness goals, such as finding an attractive mate or achieving status in the eyes of peers, results in a greater chance of success than failures to try because of a sober perspective. Considered together, the error management effects we have reviewed suggest that people should be optimistic in some circumstances but paranoid in others (i.e., they should be *paranoid optimists,* Haselton & Nettle, 2004). Whichever strategy dominates in a given situation will depend on the relative costs of errors.

In sum, biased solutions may often be better than strategies that seek to maximize accuracy. Evolutionary models of specific adaptive problems of judgment and the relative costs of errors have helped guide researchers to undiscovered adaptive biases, as well as explaining some known to exist.

Clean House and Shift the Focus

Our brief review of the errors-and-biases literature has two implications. First, now may be a good time to *clean house*. The taxonomy in Table 1 provides principled standards for deciding which biases do and do not deserve extensive study. If a bias is likely to be an artifact of a research strategy, it is unlikely to have an impact on humans' daily thoughts and actions, and surely, then, we should not devote abundant effort to studying it.

Second, a shift in focus may be in order. If errors are produced by useful heuristics that sometimes break down, they are best thought of as by-products of otherwise adaptive systems. We wonder, then, shouldn't the focus be on the adaptations themselves? Errors resulting from the use of heuristics demonstrate how the system fails, which reveals only a limited amount about its design. In investigations of personality attribution, for example, the focus has typically been on repeated demonstrations of the fundamental attribution error. The question of how observers use behavior to make reasonable inferences about enduring dispositions (a formidable task, as we will see) is neglected, leaving us with little information about how this it is actually done. Similarly, with respect to error management effects, one can investigate how and when these biases translate into adaptive social behaviors (or those that were adaptive in ancestral environments).

PERSONALITY JUDGMENT

After years of debate, most social scientists agree that personality exists—people have enduring personality traits that are useful in predicting their behavior (the only thing shocking now about this is how long it took to arrive at this conclusion; see e.g., Kenrick & Funder, 1988.) Given personality variation along important social dimensions such as cooperativeness, competitiveness, and dependability, being able to discriminate between individuals who are high and low on these dimensions poses a crucial adaptive problem. In short, when deciding with whom to cooperate and whom to avoid or whom to select as a long-term mate, better personality judgment leads to better behavioral prediction, which leads to better social decisions.

The Realistic Accuracy Model

One description of how judgment of personality characteristics might be judged accurately is the Realistic Accuracy Model (Funder, 1995, 1999). Consider an individual who has managed to accurately judge the cooperativeness, competitiveness, or conscientiousness of another. How is this possible? According to the model, four things must happen. First, the target of judgment must do something *relevant* to the trait in question. An individual cannot just sit around thinking cooperative thoughts, he or she must actually do something characteristic of cooperativeness or his or disposition will remain forever unknown. Second, this behavioral information must be *available* to the judge. A common example is physical presence. An individual might be cooperative with her family, but uncooperative at work, or vice versa, leading co-workers or family members (respectively) to underestimate her general capacity for cooperation. Third, relevant and available information must be successfully *detected* by the judge. He or she must not be so inattentive or distracted or unperceptive as to miss essential clues as to what is going on. Finally, successfully detected relevant and available information must still be correctly *utilized* by the judge, which includes being correctly remembered, compared to existing knowledge, and interpreted.

The most important implication of this model for present purposes is that it reveals how and why personality judgment is so difficult. Unless all four stages of the model are successfully traversed, accurate personality judgment is impossible, and partial imperfections at each stage combine multiplicatively (Funder, 1999). If the target does not do anything relevant, if the relevant behavior is in a context not shared with the judge (e.g., if the target deliberately conceals uncooperative behavior), if the judge misses important information, or if he or she misinterprets the information – any of these failures is sufficient to sink accuracy. Perhaps even worse, the Realistic Accuracy Model is a description of the core of the process of accurate judgment that oversimplifies the problems entailed. The model describes a one-cue one-judgment sequence, whereas in realistic contexts judges evaluate multiple traits simultaneously on the basis of multiple sources of behavioral information that vary in credibility and which derive much of their meaning from the ways in which they interact with the social context and with each other. Indeed, when one considers how difficult it is to correctly judge personality, it is possible to feel the same way some observers have felt when contemplating the formidable task of learning, comprehending and producing language: it must be impossible.

And yet, of course, people do sometimes judge personality accurately. Decades ago, Gordon Allport observed that often we *are* able "to select gifts that our friends will like, to bring together a congenial group at dinner, to choose words that will have the desired effect upon an acquaintance, or to pick a satisfactory employee, tenant, or room-mate" (Allport, 1937, p. 353). Without getting into a debate about whether people are usually right or wrong, it is easy to observe that it would be difficult to survive in a social environment if our personality judgments were not correct at least sometimes. Moreover, extensive evidence from the Riverside Accuracy Project and other sources shows that personality judgments often show impressive construct validity when evaluated against the criteria of self-other agreement, consensus, and the ability to predict future behavior (Funder, 1999).

Evidence also suggests that people can make valid inferences from very subtle appearance cues. A range of personality traits can be judged with surprising accuracy from very brief observations (Ambady, Bernieri, & Richeson, 2000), including traits associated with personality disorders (Oltmanns, Friedman, Fiedler & Turkheimer, 2004). Strangers can discriminate "cheaters" in experimental social exchange interactions from non-cheaters based on facial photographs alone (Yamagishi, Tanida, Mishima, Shimoma, & Kanizawa, 2003). Women and men tend to judge men with masculine and symmetrical faces as sexier, more sexually experienced, more dominant, but less faithful and less likely to be good dads (Johnston et al, 2001; Penton-Voak & Perrett, 2001). Research has shown that more symmetrical men (who also tend to be more facially masculine) tend toward a short-term mating strategy (Gangestad & Simpson, 2000), and thus these judgments have validity.

A Personality Judgment Instinct?

The achievement of accuracy under such difficult circumstances creates a dilemma. People seem to go way beyond the information given, to know more about personality than they should know given each individual's limited social experience, and to be better at judging attributes of others than they would be expected to be given the complex and multifaceted problem such judgment presents. A similar dilemma was resolved in the domain of language by theory and research on the "language instinct." Researchers proposed that humans are able to learn language because the infant mind already contains many language rules and specific language-learning devices (e.g., Pinker, 1994; 2000). This proposal was of course controversial, so we are aware of venturing across thin ice when suggesting a personality judgment instinct. Yet, it is worth considering the possibility that, like language, the ability to judge personality emerges during development as a result of specialized learning mechanisms. These mechanisms help developing humans to sift through cues linked with personalities by entertaining privileged hypotheses about them. In its mature form, the personality judgment instinct leads people to quickly form and utilize valid personality judgments, and thus helps to explain how accurate judgment is possible even in the face of seemingly overwhelming obstacles.

By instinct, we mean that the ability to judge personality is undergirded by a set of evolved, relatively autonomous, and specialized computational devices (or “modules”). A module, as we use the term, is similar to an organ of the body—organs are often linked and their operations can certainly affect each other, but they take in different types of bodily input and their functions are specific. Like body organs, cognitive modules might interact in some ways (Barrett, in press), but they are sensitive to only a limited range of input and they have specific functions. Fodor (1983) described many features of cognitive modules, including domain specificity, obligatory firing, rapid speed, inaccessibility to consciousness, characteristic ontogenetic course, dedicated neural structure, and a characteristic pattern of breakdown. As Fodor himself suggested, none of these features is necessary or defining (also see Barrett, in press, for an extended discussion), but observing these properties in the personality judgment system would render it unlikely that a central, general processor that is designed to achieve many different functions is responsible for personality judgment. Thus, we consider whether some of these features might exist.

Testing the Personality Judgment Instinct Theory

We have derived several hypotheses and predictions from the proposal that there is a personality judgment instinct (inspired in part by predictions by Pinker, 2000, made about language). Available evidence already supports some of these expectations, whereas others await future empirical testing.

Hypothesis 1. People should be naturally proficient in personality judgment. We have already reviewed some evidence that people form quick and generally valid personality judgments. More can be mentioned. Judgments of college professors based on 30 second silent videos predicted the professors’ evaluation scores at the end of the term (Ambady & Rosenthal, 1993). Meta-analyses demonstrate that inferences based on these thin slices of behavior are generally good (e.g., they produce moderate effect sizes in predicting criterion variables; Ambady & Rosenthal, 1992). People can even correctly judge the sexual orientation of others based on brief films of nonverbal behavior or still photographs (Ambady, Hallahan, & Conner, 1999). The literature contains many examples of this nature.

Like the parsing of one's native language or the visual recognition of objects, personality judgments should also feel effortless and natural, and we may not be able to "turn them off." Trait inferences are made without much effort, outside of awareness, and even under conditions of distraction (e.g., while simultaneously attempting to remember a long string of digits; Winter, Uleman, & Cunniff, 1985; Winter & Uleman, 1984). Indeed, it is difficult if not impossible to prevent yourself forming a first impression. If you met someone new today and shook her hand, could you prevent yourself from forming any judgments about her at all? Courtroom judges appear to understand that personality judgments are a part of human nature. They explicitly instruct juries to remain "unbiased" and to avoid jumping to conclusions about the character of the accused and the witnesses. Anyone who has served on a jury knows this is difficult if not impossible to do.

In short, people do appear to be naturally proficient. Personality judgments also seem to be characterized by three features of modules: rapid speed, inaccessibility to consciousness (at least to some degree), and obligatory firing. It is not a requirement, of course, that evolved adaptations operate in an "automatic" or nonconscious fashion. There is nothing in evolutionary theory that requires selection to design psychological adaptations that are unresponsive to contingencies involving conscious thought; and indeed, no one yet knows what the function of consciousness is (Cosmides & Tooby, 2000). Some personality judgments may be altered by deliberation and reflection and therefore the system may be cognitively penetrable. We do contend, however, that finding that some cognitive operations occur quickly and unavoidably—as appears to be the case for the parsing of language and perhaps for initial judgments of personality—suggests that they may be produced by a dedicated system.

Hypothesis 2. Personality judgment abilities should form a distinct part of the phenotype. If there is a personality judgment instinct, it should form a distinct part of the human psychological phenotype, characterized by its own specialized input conditions and decision rules, a dedicated neural structure, and subject to catastrophic breakdown (it will exist intact or its components will appear to be missing entirely).

Important components of social judgment may be missing in some individuals. People with Asperger's syndrome, a mild form of autism, have difficulty reading others' emotional expressions and understanding some subtleties of social interaction (such as when someone has committed a *faux pas*), but their other cognitive capabilities may be normal (Stone, Baron-Cohen, & Knight, 1998; Baron-Cohen, Wheelwright, Stone, & Rutherford, 1999). Similarly, damage to a region within the limbic system that is implicated in social reasoning impairs reasoning on a cheater-detection cognitive task, but not on a closely matched and logically equivalent precaution task (Stone, Cosmides, Tooby, Kroll & Knight, 2002). These two tasks are equally difficult for subjects without brain damage (Stone et al, 2002). We know of no cognitive dissociation studies specifically investigating trait inferences, but we would expect similar dissociative patterns.

Hypothesis 3. The ability to form personality judgments should emerge without explicit training and perhaps in spite of incompatible social inputs. Children begin to spontaneously use global trait-like terms and some specific trait terms as early as age 3 (Eder, 1989). Beginning at age 4, children can use trait labels like "shy" vs. "not shy" and "nice" vs. "mean" to make non-obvious inferences about mental states (Heyman & Gelman, 1999). Children from age 5 understand that two individuals who have different traits—e.g., generous vs. selfish, honest vs. dishonest—will have different emotional reactions in response to the same event (Yuill & Pearson, 1998). There is some evidence that spontaneous personality judgments also emerge early (between 3 and 6 years old) in Japan (Matsunaga, 2003).

In sum, this evidence suggests that the development of personality judgment begins spontaneously and early. Further work is needed, however, to examine whether trait inferences by children can be fully accounted for by explicit training by parents and peers. Proficiency in language use does not seem to require explicit education—people in lower class rural environments use language as complex as that of an Oxford professor. Likewise, we would expect effective personality judgment to emerge even in impoverished informational environments, and the existence of a personality judgment instinct would imply that children may develop normal abilities even if their primary social models are deficient (e.g., unaffected children of parents with autism or Asperger's syndrome).

Hypothesis 4. Personality judgment should be ubiquitous. A personality judgment instinct should be a universal part of human nature. Its behavioral manifestations may be variable, but its underlying developmental and psychological design should show evidence of universality.

One aspect of universality is that children should pass through similar developmental sequences across cultures (also see hypothesis 3). There are several clues about the developmental sequence in children from studies conducted in the United States. For example, children may first begin using general evaluative terms, then global traits, and, then specific traits (e.g., Eder, 1989; Alvarez, Ruble, & Bolger, 2001). Children everywhere may use information about motives to infer traits (Heyman & Gelman, 1998) and take hints from the lexicon about what characteristics are enduring versus transient. For example, traits picked out by nouns (*Rose is a "carrot-eater"*) result in greater attributions of stable and internal characteristics than do possible traits that are not (*Rose eats a lot of carrots*; Gelman & Heyman, 1999).

There is also evidence that personality judgments around the world converge on several personality dimensions. McCrae and Costa (1997) have amassed evidence that the big 5 structure replicates across cultures. Specific studies find that more or fewer dimensions may be needed to account for individual differences in a given culture, but 4 of the 5 factors—extraversion, agreeableness, conscientiousness, and neuroticism—appear universally robust (see Triandis & Suh, 2002, for a recent review).

Buss argued that the emergence of extraversion and agreeableness as the two major axes of individual differences reflects the universal importance for humans of discriminating others' tendencies to climb the social hierarchy and to be good partners in alliances (Buss, 1991). It seems likely that other dimensions that also had important fitness consequences will emerge consistently across cultures. Key candidates include sexual restrictedness, attractiveness (or *mate value*), health, and physical strength. We predict that people around the world will be especially proficient in forming these judgments (also see Gangestad, Simpson, DiGeronimo & Biek, 1992).

The most robust universal, however, is that people everywhere should form personality judgments. We predict that nowhere in the world will people choose mates and friends randomly with respect to personality, and even when marriages are arranged for political or financial purposes, we suspect that personality will not be irrelevant. People across the globe will also infer enduring traits in others, these inferences will generally be valid, and they will use them in making important social decisions.

Evidence of universality should also be observed within cultures. Within cultures, neurologically intact individuals who do not show evidence of specific impairment of the personality judgment system should not differ much in their abilities to form valid personality judgments, whereas they may differ widely in other abilities and preferences. A lack of variation in performance in non-verbal judgment tasks in the West may reveal that most people perform at a generally high level, perhaps near ceiling given the difficulty of the task. An enduring problem in the study of accuracy in personality judgment is that consistent individual differences in judgmental ability have been surprisingly difficult to establish (Schneider et al., 1979; Funder, 1999). Perhaps this is because personality judgment is such an essential life skill that nearly everyone can do it well enough to get by.

A further speculation is that an evolved propensity for accurate personality judgment might be particularly likely to arise for traits that have particular importance for survival and reproduction. For example, it would certainly be adaptive to be able to judge deceitfulness in one's fellow humans. But unfortunately, it would be equally adaptive to be able to feign faithfulness, so the evolutionary outcome could be a sort of "arms race" with no clear winner. By the same token, perhaps females might have a special ability to judge dominance or status (and its survival advantages for her and her offspring) in male targets, whereas males are especially sensitive to indicators of parental nurturance. But counterbalancing such possibilities is the equally likely chance that sexual selection according to these indicators might lead to their mimicry, where non-protective males would evolve misleading signs of dominance, and non-nurturing females also develop the capacity to seem other than they are. This analysis highlights the heavy task for the personality judgment instinct, which is to help people not only to detect essential attributes in others, but to see through attempts to mimic desirable traits and to mask undesirable ones.

CONCLUSIONS

Social-cognitive psychology's focus on judgmental imperfection has led research astray from focusing on the phenomenon of interpersonal judgment that truly is fundamentally important: the ability to judge personality with a useful degree of accuracy in the face of daunting obstacles. Evolution can be expected to produce a mind that produces judgments that are sufficiently accurate given cognitive and informational constraints, not perfect. Imperfection is therefore not only not surprising, it is foreordained. Many putative demonstrations of judgmental error are also artifactual. The relatively rare and especially interesting demonstrations of error are those that (1) are not merely produced by unrealistically difficult or obscurely-framed experimental tasks and (2) show misjudgments not evenly distributed around the midpoint of accuracy, but systematically biased in one direction or the other. Such findings remind us that the human cognitive system was evolved not for abstract accuracy but for survival and reproductive success. When false positives are more costly than false negatives, we should expect a bias in the direction of false negatives, and vice versa. But even judgments that are biased in this sense would still be expected to be reasonably good, as Simon pointed out long ago.

Non-artifactual findings of error and bias deserve a closer look, and a different sort of look than they have traditionally received. Rather than shaking our heads sadly at yet another demonstration of incompetence or having a “chuckle about our goofs” (Crandall, 1984, p. 1499), we should be led to ask what adaptive purpose is or was served by the cognitive system that produced these errors. Such an inquiry may lead to a deeper understanding of errors, the cognitive mechanisms that produce them, and human cognition in general.

Finally, our outline of the possible nature of a personality judgment instinct is obviously far from the final word on the matter. At present we wish to suggest that it might be heuristically useful to entertain this possibility, and to see if it helps to integrate otherwise scattered facts about personality judgment including its ease, ubiquity, universality, and general accuracy. Whether our hypotheses in the end are supported or not, evolutionarily-based analyses of interpersonal judgment are yet rare, which means the field is wide-open. Those readers of the social psychological literature who find the regular delivery from the “error of the month club” a little less thrilling than it used to be, might consider the possibilities that an evolutionarily-informed approach could offer to reinvigorate a tiring field.

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.

Allport, G.W. (1937). *Personality: A psychological interpretation.* New York: Henry Holt.

Alvarez, J. M., Ruble, D. N., & Bolger, N. (2001). Trait understanding or evaluative reasoning? an analysis of children's behavioral predictions. *Child development, 72*, 1409-1425.

Ambady, N. Bernieri, F.J., & Richeson, J.A. (2000). Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream. *Advances in Experimental Social Psychology, 32*, 201-271.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality & Social Psychology, 77*, 538-547.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256-274.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of Personality & Social Psychology*, *64*, 431-441.

Baron-Cohen, S., Wheelwright, S., Stone, V.E. & Rutherford, M.D. (1999). A mathematician, a physicist and a computer scientist with Asperger Syndrome: Performance on folk psychology and folk physics tests. *Neurocase, 5*, 475-483.

Barrett, H. C. (in press). Enzymatic computation and cognitive modularity. *Mind and Language.*

Barrett, H.C., Todd, P.M., Miller, G.F., and Blythe, P. (in press). Accurate judgments of intention from motion alone: A cross-cultural study. *Evolution and Human Behavior.*

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological bulletin, 117*(3), 497-529.

Buss, D. M. (1991). Evolutionary personality psychology. In M. R. Rosenzweig, & L. W. Porter (Eds.), *Annual review of psychology, vol. 42.* (pp. 459-491). Palo Alto: Annual Reviews, Inc.

Buss, D. M. (2003). *The evolution of desire: Strategies of human mating* (revised edition). New York: Basic Books.

Chagnon, N. (1983). Yanomamo: The fierce people (3rd ed.). New York: Reinhart & Winston.

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

Cosmides, L. & Tooby, J. (1992). Cognitive adaptations for social exchange. In Barkow, Cosmides, & Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture*. (pp. 163-228). New York: Oxford University Press.

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.

Cosmides, L. & Tooby, J. (2000). Evolutionary psychology and the emotions. In M. Lewis & J. M. Haviland-Jones (Eds.), *Handbook of Emotions, 2nd Edition*. (pp. 91-115.) NY: Guilford.

Crandall, C.S. (1984). The overcitation of examples of poor performance: Fad, fashion or fun? *American Psychologist*, *39,* 1499-1500.

Darwin, C. (1872). *The expression of the emotions in man and animals*. New York: Philosophical Library.

Eder, R. A. (1989). The emergent personologist: The structure and content of 3 1/2, 5 1/2, and 7 1/2-year-olds' concepts of themselves and other persons. *Child development, 60*(5), 1218-1228.

Eisenberger, N. I., Lieberman, M. D., & Williams, K. D. (2003). Does rejection hurt? an fMRI study of social exclusion. *Science, 302*(5643), 290-292.

Ekman, P. (1973). Universal facial expressions in emotion. *Studia Psychologica, 15*, 140-147.

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), 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, NY, England: Mcgraw-Hill Book Company.

Fodor, J. (1983). *The modularity of mind*. Cambridge: MIT Press.

Funder, D.C. (1982). On the accuracy of dispositional vs. situational attributions. *Social Cognition, 1*, 205-222.

Funder, D. C. (1987). Errors and mistakes: Evaluating the Accuracy of Social Judgment. *Psychological Bulletin, 101*, 75-90.

Funder, D.C. (1992) Everything you know is wrong (Review of *The person and the situation*). *Contemporary Psychology* 37:319-320.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*, 652-670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA, US: Academic Press, Inc.

Funder, D. C. (2003). Toward a social psychology of person judgments: Implications for person perception accuracy and self-knowledge. In J. P. Forgas, & K. D. Williams (Eds.), *Social judgments: Implicit and explicit processes; social judgments: Implicit and explicit processes.* (pp. 115-133). New York, NY, US: Cambridge University Press.

Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences*, *23*, 675-687.

Gangestad, S. W., Simpson, J. A., DiGeronimo, K., & Biek, M. (1992). Differential accuracy in person perception across traits: Examination of a functional hypothesis. *Journal of Personality & Social Psychology, 62*, 688-698.

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories. *Psychological Science, 10*(6), 489-493.

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

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

Gigerenzer, G., Czerlinski, J., & Martignon, L. (2002). How good are fast and frugal heuristics? In T. Gilovich, & D. Griffin (Eds.), *Heuristics and biases: The psychology of intuitive* *judgment; heuristics and biases: The psychology of intuitive judgment.* (pp. 559-581). New York: Cambridge University Press.

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

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

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

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

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

Haselton, M. G. & Buss, D. M. (2003). Biases in Social Judgment: Design Flaws or Design Features? In J. Forgas, K. Williams, & B. von Hippel (Eds.) *Responding to the social world: implicit and explicit processes in social judgments and decisions.* New York, NY: Cambridge.

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

Haselton, M. G., Nettle, D. & Andrews, P.W. (in press). The evolution of cognitive bias. In D. M. Buss (Ed.), *The Evolutionary Psychology Handbook*, Wiley.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243-259.

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

Heyman, G. D., & Gelman, S. A. (1998). Young children use motive information to make trait inferences. *Developmental psychology*, *34*, 310-321.

Heyman, G. D., & Gelman, S. A. (1999). The use of trait labels in making psychological inferences. *Child development*, 70, 604-619.

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

Hrdy, S. (1999). *Mother nature: A history of mothers infants, and natural selection.* New York: Pantheon Books.

James, W. (1890). *The principles of psychology* (2nd ed). Chicago: Encyclopedia Britannica (published in 1990).

Johnson-Laird, P. N., Legrenzi, P., & Legrenzi, M. S. (1972). Reasoning and a sense of reality. *British Journal of Psychology, 63*(3), 395-400.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone mediated adaptive design. *Evolution and Human Behavior, 21*, 251-267.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist, 58,* 697-720.

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

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

Kenrick, D.T., & Funder, D.C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*, 23-34.

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

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

Kurzban, R. & Aktipis, A. (in press). Modular minds, multiple motives. To appear in Schaller, M., Simpson, J., & Kenrick, D. (Eds.) *Evolution and Social Psychology.*

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

Kyl-Heku, L. M., & Buss, D. M. (1996). Tactics as units of analysis in personality psychology: An illustration using tactics of hierarchy negotiation. *Personality & Individual Differences, 21,* 497-517.

Lopes, L. L. (1991). The Rhetoric of Irrationality. *Theory and Psychology, 1*, 65-82.

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

Matsunaga, A. (2002). Preschool children's inferences about personality traits. *Japanese Journal of Developmental Psychology, 13,* 168-177.

McCrae, R. R., & Costa, P. T. J. (1997). Personality trait structure as a human universal. *American Psychologist, 52,* 509-516.

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

Miller, G. F. (2000). *The mating mind*. New York: Doubleday.

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

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

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

Oltmanns, T.F., Friedman, J.N.W., Fiedler, E.R., & Turkheimer, E. (2004). Perceptions of people with personality disorders based on thin slices of behavior. *Journal of Research in Personality, 38,* 216-229.

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

Park, J. H., Schaller, M., and Crandall, C. S. (2004). Obesity as a heuristic cue connoting contagion: Perceived vulnerability to disease promotes anti-fat attitudes. Manuscript under review.

Penton-Voak, I. S. & Perrett, D. I. (2001). Male facial attractiveness: Perceived personality and shifting female preferences for male traits across the menstrual cycle. *Advances in the Study of Behavior, 30,* 219-259.

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

Pinker, S. (1994). *The language instinct*. New York, NY, US: William Morrow & Co, Inc.

Pinker, S. (2000). Language as an adaptation to the cognitive niche. In M. Christiansen & S. Kirby (Eds.), *Language evolution: Report from the research frontier*. New York: Oxford.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 3*, 781-799.

Ross, L. (1977) The intuitive psychologist and his shortcomings. *Advances in Experimental Social Psychology* 10:174-214.

Schaller, M., Park, J. H., & Faulkner, J. (2003). Prehistoric dangers and contemporary prejudices. *European Review of Social Psychology, 14*, 105-137.

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.

Schmitt, D. P., & International Sexuality Description Project. (2003). Universal sex differences in the desire for sexual variety: Tests from 52 nations, 6 continents, and 13 islands. *Journal of Personality & Social Psychology, 85*, 85-104.

Schneider, D.J., Hastorf, A.H. & Ellsworth, P.E. (1979). *Person perception*. Reading, MA: Addison-Wesley.

Simon, H.A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*, 129-138.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*, 640-656.

Stone, V. E., Cosmides, L., Tooby, J., Kroll, N. & Knight, R. T. (2002). Selective impairment of reasoning about social exchange in a patient with bilateral limbic system damage. *Proceedings of the National Academy of Sciences, USA. 99,* 11531-11536.

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. (1988). *The evolution of war and its cognitive foundations*. Institute for Evolutionary Studies Technical Report #88-1.

Tooby, J. & Cosmides, L. (1996). Friendship and the Banker's Paradox: Other pathways to the evolution of adaptations for altruism. In W. G. Runciman, J. Maynard Smith, & R. I. M. Dunbar (Eds.), Evolution of Social Behaviour Patterns in Primates and Man. *Proceedings of the British Academy, 88*, 119-143.

Tooby, J. & DeVore, I. (1987). The reconstruction of hominid behavioral evolution through strategic modeling. In W. Kinzey (ed.) *Primate Models of Hominid Behavior*. New York: SUNY Press.

Triandis, H. C., & Suh, E. M. (2002). Cultural influences on personality. *Annual Review of Psychology, 53*(1), 133-160.

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.

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

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

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.

Williams, K. D., Cheung, C. K. T., & Choi, W. (2000). Cyberostracism: Effects of being ignored over the internet. *Journal of Personality & Social Psychology, 79*(5), 748-762.

Williams, K. D., & Sommer, K. L. (1997). Social ostracism by one's coworkers: Does rejection lead to loafing or compensation? *Personality and Social Psychology Bulletin, 23*, 693-706

Windschitl, P. D., & Chambers, J. R. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 30*(1), 198-215.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality & Social Psychology, 47*, 237-252.

Winter, L., Uleman, J. S., & Cunniff, C. (1985). How automatic are social judgments? *Journal of Personality & Social Psychology, 49*, 904-917.

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E., & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution & Human Behavior, 24*(4), 290-301.

Yuill, N., & Pearson, A. (1998). The development of bases for trait attribution: Children's understanding of traits as causal mechanisms based on desire. *Developmental psychology, 34(3),* 574-586.

Table 1: Evolutionary Taxonomy of Evidence of Bias and Error in Social Psychological Research

Cause of Apparent Bias	Examples
Artifact: Apparent biases and errors are artifacts of research strategies. Biases result from the application of inappropriate norms (e.g., Cosmides & Tooby, 1996), the placement of humans in unnatural settings (e.g., Gigerenzer, 1997), or testing within the *error paradigm* (Krueger & Funder, in press)	• Some instances of base-rate neglect (Hertwig & Gigerenzer, 1999) • Some instances of confirmation bias (Cosmides, 1989)
Heuristic: Bias results from the use of heuristics, which work well in most circumstances but are prone to systematic breakdown. Heuristics are compromise solutions to problems of judgment given time or processing capacity constraints (e.g., Tversky & , Kahneman, 1974) or ecological/informational constraints (e.g., Gigerenzer, Czerlinski, & Martington, 2002).	• Fundamental Attribution Error • One-reason Decision Strategies (Gigerenzer, Todd, et al., 1999)
Error Management: Selection favors bias toward the less costly error (Haselton & Buss, 2000). Error management causes overall rates of error to increase, though net costs are minimized.	• Auditory Looming (Neuhoff, 2001) • Sexual Overperception by Men (Haselton, 2003) • Commitment Skepticism in Women (Haselton & Buss, 2000)