In D. A. Hofmann & M. Frese (Eds.),
Errors in organizations (pp. 97–112). New York: Taylor & Francis. (2011)

4

Revisiting the "Error" in Studies of Cognitive Errors

Shabnam Mousavi and Gerd Gigerenzer

A few decades ago, cognitive scientists studied judgment errors to discover rules that govern our minds, just as visual errors were studied to unravel the laws of perception. This practice has generated a long list of so-called cognitive biases, with disappointingly little insight into how the human mind works. In this chapter, we present our diagnosis of this fruitless labor. An important difference between errors of perception and those of judgment is that visual perception is measured against properties that are objectively measurable, whereas judgments are traditionally evaluated against conventional norms for reasoning and rationality. Moreover, most visual perception errors have developed evolutionarily and are "good" in the sense of enhancing survival or adaptation to the natural or habitual environment, whereas judgmental errors are traditionally perceived as biases that encumber decision making. We maintain that humans can and do make both "bad" and "good" errors. Bad errors, such as spending more than what you earn, should be avoided, whereas good errors are valuable. An example of good errors is the grammatical mistakes that children make with irregular verbs, such as "I *breaked* my toy!" It is through continuous feedback to such errors that children master their native language (for a discussion of good errors, see Gigerenzer, 2005). Many innovations are results of playfulness, allowing for errors and learning from these. One celebrated failure in industry is the accidental development of a weak glue (instead of a strong one), which led to the invention of the Post-it™.* Good errors are side effects of reasonable inferences, explanations, or even serendipity.

Human judgments are usually considered erroneous when measured against logical and statistical norms of rationality. Hence, we address two essential questions of cognitive studies: What is rational judgment? How can one construct reasonable cognitive norms? To answer these questions, we scrutinize two phenomena famous for supposedly demonstrating

* We are thankful to Reza Kheirandish for providing us this illustrative example.

human logical and calculative errors: the Wason selection task and overconfidence. First, we show that the Wason selection task confuses logic with rationality, specifically social rationality. Second, we distinguish five types of overconfidence and the environmental structures in which they appear. These two steps bring to light our view of rational judgment and proper norms: Rational judgment must be evaluated against an ecological notion of rationality, which in turn requires constructing content-sensitive norms. In contrast to logical norms, which are content blind in assuming the truth of syntax, content-sensitive norms reflect the actual goals and specifics of the situation. Ecological rationality is about the match of decision-making strategies to the structure of information in the environment. This match is an ecological one but not a mirror image of the environment. Finally, while the literature on cognition claims that "debiasing" is hardly possible, we illustrate the contrary.

Finally, humans react to different representations of information differently. Thus, the way in which information is presented to decision makers can enhance or hinder sound judgment. Using examples from the medical field, in which patients and doctors have to make vital decisions under pressures of time, emotions, and money, we demonstrate how communication of information can be enhanced through transparent modes of presenting risk information.

Is a Violation of the Logical Truth Table an Error in Judgment?

One of the most studied reasoning problems, the Wason (1966) selection task (a Google search returned more than 150,000 entries), assumes logic as a norm for evaluating people's judgments and supposedly demonstrates logical errors.

Various Wason selection tasks share the same logical structure. In the four-card game, the following rule holds: *If P then Q*. The cards shown have information about four situations, with each card representing a situation. One side of a card tells whether *P* has happened, and the other side of the card tells whether *Q* has happened. The player needs to indicate only those cards that definitely need to be turned over to verify whether any of the cards violate the rule.

P	not-*P*	*Q*	not-*Q*

The answer that satisfies logical truth is to choose the *P* card (to see if there is a not-*Q* on the back) and the not-*Q* card (to see if there is a *P* on the back). This is because the rule is violated by any situation in which *P* happens and *Q* does not. But, this logic does not constitute a useful norm for all kinds of reasoning. For instance, let us apply it to the task of cheating detection. We will see that pragmatic goals affect judgment in ways that are not necessarily compatible with logical truth yet cannot be labeled as errors.

Consider the following conditional: "If an employee works on the weekend (*P*), then that person gets a day off during the week (*Q*)." In an experiment (Gigerenzer & Hug, 1992), two roles were assigned to participants: employee or employer. Both were given the same task: to detect whether the rule has been violated. For those who played the role of employee, their dominant response coincided with the logically correct answer, *P* and not-*Q*. However, for those who played the role of employer, the dominant response changed to not-*P* and *Q*. Why? The answer lies in the perspective of players, who define cheating according to their assigned roles. For the employee, the rule has been violated (i.e., cheating has occurred) if weekend work has been performed (*P*) but not rewarded with a day off (not-*Q*). To determine this, the employee needs to inspect *P* and not-*Q*, which coincides with the logically correct verification of the conditional. Therefore, in this particular case, the goal of verifying logical truth and that of detecting cheaters lead to the same pattern of information search. From the perspective of employers, however, cheating has happened when an employee takes a weekday off (*Q*) without earning it through weekend work (not-*P*). This explains why employers would select not-*P* and *Q* instead of trying to verify the logical truth that requires selecting *P* and not-*Q*. When logical truth is used as a benchmark, half of the players appear to be following it (the employees' action), whereas the other half do not. That is, using logical truth as a measure of erroneous choice is irrelevant. The achievement of social contract theory (Cosmides, 1989) was to replace the logical rule with the (evolutionary) motivation of cheating detection, thus predicting both observed outcomes. Gigerenzer and Hug (1992) took another step forward by decomposing the observed behavior into correspondence with cheating detection and obligations of a social contract (which defines roles and sets goals.). They showed that "it is *the pragmatics* of who reasons from what perspective to what end" that can sufficiently account for the observed behavior.

Proper norms are sensitive to pragmatics and semantics. We are against upholding logical and statistical rules as a universal yardstick for evaluating human judgment; such "content-blind" norms misrepresent the actual goals (pragmatics) of action. Notice, for example, that in the case of employee/employer role-playing, researchers who held logical truth to be the only yardstick for measuring human rationality evaluated some judgments as rational (e.g., when participants were the employee) and others as irrational (when participants were the employer). In contrast, a content-sensitive approach would predict and explain both patterns of judgments by the adaptive (not Kantian) goal of detecting whether the *other* party has cheated, which is a necessary part of social exchange, social contract, and forms of reciprocal altruism (Trivers, 2002).

An interesting result from the experiments reported in Gigerenzer and Hug (1992) is that some participating students remained true to their training in logic, ignoring the content and always looking for the logically correct answer. However, when interviewed, these subjects reported that what they chose according to logical norms "doesn't make sense."

Does this mean that we should abandon logic? No. Should we abandon logic as a universal, content-blind norm for rational thinking, as assumed by psychologists such as in Wason's (1966) four-card task and Tversky and Kahneman's (1983) conjunction fallacy problems? Yes. To establish a reasonable and useful theory of norms, we need to start with the actual goals of humans, not with the goal of maintaining logical truth. As shown again in the following section, many a puzzling "bias" resolves itself if the focus shifts to content-sensitive norms. We disagree with the view that "our minds are not built (for whatever reason) to work by the rules of probability" (Gould, 1992, p. 469). Rather, biases often appear when researchers ignore the semantics, assume the logical truth of syntax, and reduce the pragmatics of action to fit the observed behavior into convenient and conventional analytical frameworks. These are the systematic errors *committed by researchers*. This is not to say that ordinary people do not make bad errors, but that researchers should reason more carefully about good reasoning in the first place. This argument applies to the celebrated "overconfidence bias" as well.

Is "Overconfidence Bias" an Error?

"The most robust effect of overconfidence" (Odean, 1998, p. 1888) was elaborated in the framework of the research program generally referred to as *heuristics and biases* (Kahneman, Slovic, & Tversky, 1982). A continuing source of confusion within this program is the fact that different phenomena have been labeled with the same word. For instance, the labels "availability" and "representativeness heuristic" have been used for at least a dozen different phenomena, including mutually exclusive ones (Ayton & Fischer, 2004; Gigerenzer, 1996, 2006). The same applies to "overconfidence bias." In the following, we argue that there are at least five different phenomena labeled as overconfidence and its close cousin, (excessive) optimism. We then clarify the problematic nature of imposing improper standards as a benchmark for the evaluation of human judgment.

Overconfidence 1: Better Than Average

Svenson, Fischhoff, and MacGregor (1985) report "an optimism bias: a tendency to judge oneself as safer and more skillful than the average driver, with a smaller risk of getting involved and injured in an accident" (p. 119). Their study is a logical continuation of earlier studies (see also Svenson, 1978, and Johansson & Rumar, 1968). For example, 77% (in Svenson, 1981) of the subjects responded "yes" to the question of whether they drive more safely than the average driver. The authors explain: "It is no more possible for most people to be safer [drivers] than average than it is for most to have above average intelligence" (Svenson et al., 1985, p. 119). Yet, this claim is correct only if both safe driving and intelligence have distributions that are symmetric rather than skewed. In reality, the distribution of car accidents is asymmetric—unlike IQ distributions, which are standardized to be normally distributed and symmetric so that the median and mean are the same. The asymmetric (skewed) distribution of car accident data implies that more than 50% of drivers do in fact have fewer accidents than the mean. To check the size of the effect, we analyzed the data on approximately 7,800 drivers in the United States, reported by Finkelstein and Levin (2001). Among these American drivers, 80% indeed had fewer than the mean number of accidents.

This oversight of the difference between symmetric and asymmetric distributions can be counterbalanced by asking for percentiles or medians rather than averages or means and is sometimes done. Yet, asymmetry is not the only problem with the claim that people commit reasoning errors.

The second problem with the normative argument is that people are asked ambiguous questions for which correct answers cannot be determined. The question of whether one is an "above average" driver or an "above average" teacher leaves open to what "average" refers, and the statement will be interpreted in individual ways. For some, a good driver is one who causes no traffic accidents; for others, it is someone who can drive faster than others, hold a mobile phone and a cigarette while driving, or obey the law and drive defensively. The subjects who are asked to rate themselves on driving safety in comparison with the average driver will likely assume their understanding of a good driver. If people tend to be better than others at what they believe one should do, the ambiguous question is a second, independent reason to produce the "better-than-average" effect.

There is also a third reason why researchers need to be cautious with quick judgments about what is right and wrong. Consider the question, "Is your IQ above or below the mean of 100?" Such a question avoids the asymmetry argument and to some degree the second argument as well because it is fairly precise (although different IQ tests can lead to strikingly different results for the same person). Even here, the normative claim that "50% of all people have an IQ over 100" is not strictly correct. The "Flynn effect" implies that IQ climbs annually on average around 0.3 points, that is, there is a "year effect" associated with IQ measurements. Therefore, the statistical tests must be restandardized on a regular basis. Yet, the more years that pass after the last standardization, the more people will have above-average IQs. In these cases, the statement that more than 50% of people are better than average is, again, a fact about the world, not a judgment error.

The general point here is that to make statements about people's irrationality, researchers need to analyze the statistical structure of the environment carefully, pose clear questions, and take into account the reference points that people use for making their judgments. For the study on self-assessments of driving skills, this includes verifying the distribution of car accidents and what a person considers good driving. Only then can one decide whether a person's judgment is based on the kind of self-deception that the term *overconfidence* suggests or whether the problem lies mostly in habitual normative beliefs inside researchers' minds.

Overconfidence 2: (Too) Narrow Confidence Intervals

A second phenomenon also called overconfidence is that individuals tend to indicate too narrow confidence intervals. For instance, subjects of an experiment were asked to guess an interval for the next year's interest rate that included the true value with a probability of .90. The true value, however, was only included in 40 to 50% of the intervals (Block & Harper, 1991; Lichtenstein, Fischhoff, & Philips, 1982; Russo & Shoemaker, 1992). Because the intervals produced were, on average, too small, the conclusion was drawn that this is another case of "oversimplification." Note first that Overconfidence 1 and 2, in spite of sharing the same name, are distinct phenomena. Better than average (Overconfidence 1) does not imply narrow confidence intervals (Overconfidence 2) or vice versa. We do not know of a single study demonstrating that people who show the first phenomenon also show the second. Labeling different phenomena as the same without any evidence is one thing; calling a phenomenon an error is another.

Again, important insight would be gained by carefully specifying reasonable norms and models (rather than labels) of the cognitive processes that imply the observed phenomenon (see Juslin, Wennerholm, & Olsson, 1999; Winman, Hansson, & Juslin, 2004). Juslin, Winman, and Hansson (2007) presented a naïve sampling model, which offers a causal explanation for Overconfidence 2. This model is based on the fact that if everyday experiences (random samples) are transferred directly to a population, statistical theory implies that a systematic error will appear in estimating the variance and therefore in producing confidence intervals. Estimation of means, however, remains unaffected in this situation. Juslin et al.'s naïve sampling model enables precise predictions of produced confidence intervals and was successfully tested in a series of experiments. If the experimenter provides a probability (an average) and asks about an interval for this probability, the reported phenomenon of Overconfidence 2 appears. However, if an interval is provided and subjects are asked to provide the probability that the true value lies in the interval, the same phenomenon disappears. Juslin et al.'s scrutiny illustrates how this phenomenon (Overconfidence 2) can be elicited or eliminated from the same subject according to the setup and therefore cannot represent a stable personality characteristic or trait. Overconfidence 2 can be adequately explained by the fact that means are unbiased, but intervals (variance) are biased estimators.

Overconfidence 3: Mean Confidence Is Greater Than Percentage Correct

There is a third phenomenon also called overconfidence. It initially emerged from answers obtained to two-step questions such as "Which city lies further south: Rome or New York? How certain are you that your answer is correct?" Since the 1970s, many studies of this sort have shown that people express a level of certainty (confidence) in their own judgment that is higher than the average percentage of correct answers. The result was labeled "overconfidence" and put into the same category as the two previously mentioned phenomena, suggesting more evidence for the same irrational propensity. In one of these studies, Griffin and Tversky (1992) asserted, "The significance of overconfidence to the conduct of human affairs can hardly be overstated" (p. 414).

The probabilistic mental models theory (Gigerenzer, Hoffrage, & Kleinbölting, 1991) provided a model (rather than a label) for confidence judgments and predicted for the first time that Overconfidence 3 appears if the experimenter systematically selects the tasks and disappears if tasks are chosen randomly. These results were experimentally confirmed. As a rule, judgments were well adapted to reality but erred when participants were faced with selected and therefore nonrepresentative tasks. For example, the annual average temperature of a city is generally a good cue for its geographic location. When people are asked whether Rome or New York lies further south, Overconfidence 3 appears because knowledge about the annual average temperature of the two cities leads to an incorrect answer: Many do not know that this selected pair is an exception to the rule, and that Rome is actually higher in latitude than New York.

Gigerenzer et al. (1991) attributed the emergence and disappearance of overconfidence to selecting questions of this type versus using a representative set. This finding was first denied (Griffin & Tversky, 1992) but later supported in an extensive analysis of 130 studies, in which this form of overconfidence appeared only if tasks were selected systematically rather than randomly (Juslin et al., 2000).

Just as the naïve sampling model reveals the mechanisms of overconfidence in the case of seemingly narrow confidence intervals (Overconfidence 2), the probabilistic mental models theory reveals the phenomenon of overconfidence to be a mismatch between a mental model adapted to an environment and experimenters' biased task selection (Overconfidence 3). These theories explain and predict when Overconfidence 2 and 3 appear or disappear in the same person. The fact that the appearance or disappearance of Overconfidence 3 is subject to whether questions are sampled in a selective or representative way again demonstrates that the phenomenon is a poor candidate for a human trait or personality type, as the label suggests. Nonetheless, many researchers persist in asserting that overconfidence is a general trait or disposition.

Overconfidence 4: Overconfidence Equals Miscalibration

A fourth definition of overconfidence is miscalibration. Here, data are elicited from questions similar to those in the previous case (Overconfidence 3): "Which city lies further south: Rome or New York? How certain are you that your answer is correct?" What is analyzed, however, is not the discrepancy between the average confidence and the proportion correct, but rather the total calibration curve. For instance, a typical finding is that in all cases when people said they were 100% sure that their answer was correct, the proportion correct was only 80%; when people said that they were 90% confident, the proportion correct was only about 70%; and so on. Here, calibration is assumed to mean that the confidence categories match the proportion correct, and the mismatch is called miscalibration, or overconfidence. This again is a distinct phenomenon that can appear independently from Overconfidence 3. For example, although the difference between mean correct and percentage correct (Overconfidence 3) can be zero (e.g., when the curve for the proportion correct crosses the calibration line at 50%, as in a regression curve), miscalibration can be substantial.

For two decades, the fact that the actual curve of proportion correct differed from the calibration line was attributed to deficits of the human brain until it was noticed independently by Erev, Wallsten, and Budescu (1994) and Pfeifer (1994) that researchers had made an error: They had overlooked that a regression to the mean is at work. Confidence judgments tend to generate noisy data—that is, conditional variance is larger than zero, which is equivalent to assuming that the correlation between confidence and proportion correct is imperfect. An imperfect correlation implies that when the reported confidence ratings are high, the corresponding proportions correct will be smaller and hence resemble miscalibration and overconfidence. Typically, for general knowledge questions sampled randomly from a large domain, the regression line is symmetrical around the midpoint of the reported confidence scale (e.g., a midpoint of 50% when the confidence scale is from 0 to 100% and 75% when the confidence scale is from 50% to 100%; Juslin et al., 2000). That is, to unskilled eyes the mere presence of conditional variance can appear as systematic bias in the judgments, where there is only unsystematic noise—just as in Francis Galton's famous example that sons of tall fathers are likely to be smaller in height, and sons of small fathers are likely to be taller (see Stigler, 2002, for a detailed account). It is a normal consequence of regression, not a cognitive bias. In these environments, any intelligent system, human or computer, will produce patterns that mimic what has been called miscalibration or overconfidence.

If one estimates the confidence judgments from proportion correct (rather than vice versa), regression implies the mirror result: a pattern that looks as if there were underconfidence bias. When looking at all items that the participants got 100% correct, for instance, one will find that the average confidence was lower, such as 80%. This seems to indicate underconfidence. In contrast, when looking at all items for which participants were 100% confident, one finds that the average proportion correct was lower, such as 80%. This seems to indicate overconfidence. Erev et al. (1994) showed for three empirical data sets that regression toward the mean accounted for practically all the effects that would otherwise have been attributed to overconfidence or underconfidence, depending on how one plotted the data. Dawes and Mulford (1996, p. 210) reached the same conclusion for another empirical data set. In general, one could determine whether there is under-/overconfidence beyond regression by plotting the data both ways. This research shows that Overconfidence 4 appears largely due to researchers' misinterpreting regression to the mean as a cognitive illusion of their experimental subjects.

Overconfidence 5: Functional Overconfidence

The fifth and last phenomenon that is labeled overconfidence is of yet another nature. In this case, a functional and often profitable mechanism is at play. Many tasks, such as predicting the stock market, are so difficult that even experts do not perform better than laypeople, and sophisticated statistical strategies are not consistently better than intuition. For instance, in a comparison of a dozen optimization methods and the simple $1/N$ heuristic (equal allocation of money to all options) for investment allocation decisions, $1/N$ proved to be as good as or better than complex statistical methods in most cases (DeMiguel, Garlappi, & Uppal, 2009). In other studies, professional analysts performed worse than chance in picking stocks, while laypeople performed at chance level. Nevertheless, many customers would like to believe that experts can predict the stock market, or they tend to delegate responsibility for selecting strategies to experts. In such situations, experts who convey security and confidence stand a good chance of winning trust even if it is ill founded. In medicine, placebo effects are based on this kind of belief in the efficacy of treatments that have no real effect. Without unrealistic self-confidence, option advisers and astrologers would lose their customers—and some physicians their patients. Instances of functional overconfidence give us an opportunity to study the creation of the illusion of certainty and its role in determining human behavior (Gigerenzer, 2002). Functional overconfidence has little to do with the previously listed four types of overconfidence.

The story of the overconfidence bias may tell us more about the biased norms of many researchers in this field than the biased thinking of their experimental subjects. We pointed to two flaws: First, the practice of referring to different phenomena using the same label impedes understanding their nature; second, the norms against which people's judgments have been evaluated as flawed are often flawed themselves (see Gigerenzer, 1996, 2000). Moreover, deviations of human judgment measured against these standards are not predicted by formal psychological theories but are instead explained away by vague "irrational" notions such as overconfidence. Functional overconfidence, however, illustrates that there are parts of our external world where too much certainty is expected, and where judgments are thus directed by other incentives and goals than factual correctness.

Overconfidence research should be of particular interest to people who study the lack of progress made by psychological theories. Even after the normative problems had been pointed out in top journals such as *Psychological Review* (Gigerenzer et al., 1991; Juslin et al., 2000), many researchers nonetheless maintained that there is clear evidence for a general human tendency toward overconfidence, and one can still hear this message in social psychology, behavioral economics, and behavioral finance. Indeed, there is one clear instantiation of this bias: *the overconfidence of many researchers in a phenomenon called overconfidence.*

One of the dangers of focusing on toy problems such as the Wason selection task and many overconfidence tasks is that researchers lose sight of the real errors that people make in the world outside the laboratory. In the next section, we illustrate how an ecological rather than a logical approach to cognition can help people to learn how to reduce these real-world errors.

Helping People Avoid Judgment Errors

In this section, we present some findings from the study of decision-making processes in the health care arena on how doctors and patients make decisions. The health statistics that doctors obtain and communicate to their patients can be reported in different formats, such as relative risk (percentage change), absolute risk (how many in what total), or conditional probabilities, including sensitivity and false-positive rate. In what follows, we explain these modes of representation and recommend particular representations that reduce the potential of confusing health providers and decision makers.*

Absolute Risk Versus Relative Risk

"Mammography screening reduces the risk of dying from breast cancer by 20%" is an example of reporting results of a clinical study in terms of relative risk. In this form, the baseline information is concealed; we do not know "20% of how many." An absolute risk, in contrast, reveals the baseline risk. The absolute risk reduction is the absolute difference between the treatment and control groups. For instance, the same result can be reported in absolute risk terms as "Mammography screening reduces the risk of dying from breast cancer by about 1 in 1,000, from about 5 in 1,000 to about 4 in 1,000." That is, the 20% corresponds to 1 in 1,000. In a study with 160 gynecologists, one third did not know that the relative risk of 20% (sometimes reported as 25%) means 1 in 1,000 but instead believed that it means 20 or 200 in 1,000 (Gigerenzer et al., 2007). To communicate risk reduction associated with treatments, we therefore recommend using absolute risks, not relative risks.

Conditional Probability versus Natural Frequency

Prevalence, sensitivity, and false-positive rate are the usual pieces of information that physicians need when assessing positive predictive values, that is, the probability that a patient has a disease given a positive screening test result. Assume that in a particular region, the prevalence of the disease (probability of a woman having breast cancer) is 1%, sensitivity is 90% (probability of a woman with breast cancer testing positive), and the false-positive rate is 9% (i.e., probability of a woman without breast cancer testing positive). What then are the chances that a woman who has tested positive actually has breast cancer? When this information was provided

* This section heavily draws on the work of Gigerenzer, Gaissmaier, Kurz-Milke, Schwartz, and Woloshin, 2007.

to a group of gynecologists, their answers varied between a 1% and a 90% chance of cancer, with a majority of them overestimating the chances to be above 80%. However, when these physicians were given an alternative representation of the same information involving natural frequencies, their judgments improved dramatically, such that 87% of them correctly specified the chances of cancer for a woman with a positive test result to be about 1 in 10.

The first mode of presentation uses a *conditional probability* format, which is the probability of an Event A given an Event B (such as probability of testing positive if having cancer). Because of the difficulties it causes the majority of physicians in assessing the patient's chances of cancer, we consider it to be a nontransparent mode of risk communication. The alternative mode, called *natural frequency* format, communicates the same information in terms of raw counts for each group out of a certain total, as the following example shows: Of 1,000 women, we expect that 10 have breast cancer; of these 10 women with cancer, 9 test positive; and of 990 women without cancer, about 89 still test positive. Thus, we expect that only 9 out of 98 who test positive actually have cancer. Our recommendation is to use natural frequencies and avoid conditional probabilities.

In sum, whether physicians and patients make errors in estimating probabilities depends heavily on the way information is presented in the environment. We also believe that human choices can be enhanced—for instance, mistakes (or avoidable errors) can be reduced—by designing environments that facilitate the access to information (Gigerenzer & Hoffrage, 1995; Hertwig & Gigerenzer, 1999). As a guideline, we recommend that the questions given next always be asked about all types of risk information. The answers to these questions clarify the numbers and improve the reliability and correctness of the information communicated. These questions apply to any situation of decision making under risk; here, the explanations following each question are taken from the medical field.

1. "Risk of what?" Is it risk of getting a disease, developing a symptom, or dying from the disease?
2. "Time frame?" Risk changes over time and directly affects the evaluation of treatment options. So, it is important to present and receive information for specific time frames, such as "over the next 10 years," as opposed to "over a lifetime."
3. "How big?" The size of risk is best communicated in absolute terms, for example, 2 of 10,000 instead of a percentage with an unclear baseline. In addition, the associated risk attached to a certain condition (such as risk of cancer for smokers) can be compared to tangible fatal events (such as car accidents).

4. "Does it apply to me?" For the risk analysis available from a study to be relevant to you, the study group should include people like you. This means that you must share determining characteristics with this group. For example, if the age group of people in the study is different from yours and age is a significant determinant of the studied disease, the risk results are unlikely to apply to you. Recall that Overconfidence 2 was explained as the result of extending results from a random sample to a population. Keeping this in mind, you can avoid making the mistake of extending and using results that are found for a sample to which you do not belong.

An Ecological View of Error

In the study of cognitive errors, the tradition of focusing on logic as a universal yardstick for evaluating judgments (which produces a list of irrationalities) is an unfruitful practice because it measures deviations from content-blind norms. Content-blind norms accept logical, mathematical, or statistical truths at face value and implement them as such in models of decision making. However, generalizing from the world of models to the real world is often unjustified or inadequate. As shown in our close examinations of two popular "errors," namely, violations of logic in the Wason selection task and the overconfidence bias, content-sensitive norms are needed to evaluate behavior in the real world. Sensible norms can be conceived in an ecological framework, in which many biases disappear as soon as the actual structure of information in the environment is taken into account and the focus is turned to finding an *effective* match between the environment and available strategies. This ecological view of judgment and its resulting methods—for instance, utilizing favorable modes of information presentation—can enhance human performance.

The framework presented in this chapter does not entail exaggerated assumptions about cognitive abilities or oversimplification of the environment. In this ecological view of error, people of course make mistakes. As outlined, there are three types of errors: good errors, bad errors, and nonerrors. Some errors are useful and need be recognized as such (e.g., grammatical mistakes in the process of learning a language). Other errors are bad and should be avoided (e.g., misunderstanding medical test results). Many bad errors can be avoided by improving presentations of information, using forms that are easily absorbed and interpreted by human minds (e.g., use of absolute risks for communicating risk). Finally, some phenomena (e.g., overconfidence) that are mislabeled as errors can be easily explained if scientists include more than the conventional norms in their operational frameworks and allow for a wider scope of analysis.

Central to this chapter is the argument that the study of cognitive errors has not been successful in unraveling rules of human mind mainly because it focuses on abstract, incomplete, and sometimes irrelevant norms. One might ask, "Why does it matter if a scientific inquiry is focused on erroneous questions for a while? We can still learn *something*!" In light of what was presented in this chapter, we hope that you agree that it is time to rectify the questions and move forward. Allow us to present a strategy for doing so: to look beyond the most convenient analytical tools and rethink our habits of scientific practice. We hope that the current piece encourages less-biased practice of scientific inquiry into unraveling processes of the human mind.

References

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: Two aces of subjective randomness? *Memory & Cognition, 32*, 1369–1378.

Block, R. A., & Harper, D. R. (1991). Overconfidence in estimation: Testing the anchoring and adjustment hypothesis. *Organizational Behavior and Human Decision Processes, 49*, 188–207.

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.

Dawes, R. M., & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment, or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes, 65*, 201–211.

DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/*N* portfolio strategy? *Review of Financial Studies, 22*, 1915–1953.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review, 101*, 519–527.

Finkelstein, M. O., & Levin, B. (2001). *Statistics for lawyers* (2nd ed.). New York: Springer-Verlag.

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

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

Gigerenzer, G. (2002). *Calculated risks: How to know when numbers deceive you*. New York: Simon & Schuster. (UK edition: *Reckoning with risk: Learning to live with uncertainty*. London: Penguin Books, 2002)

Gigerenzer, G. (2005). I think, therefore I err. *Social Research, 72*, 1, 195–218.

Gigerenzer, G. (2006). Heuristics. In G. Gigerenzer & C. Engel (Eds.), *Heuristics and the law* (pp. 17–44). Cambridge, MA: MIT Press.

Gigerenzer, G., Gaissmaier, W., Kurz-Milke, E., Schwartz, L., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*(2), 53–96.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*, 506–528.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*, 684–704.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43*, 127–171.

Gould, S. J. (1992). *Bully for brontosaurus: Further reflections in natural history*. London: Penguin Books.

Griffin, D. W., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411–435.

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

Johansson, G., & Rumar, K. (1968). Visible distances and safe approach speeds for night driving. *Ergonomics, 11*, 275–282.

Juslin, P., Wennerholm, P., & Olsson, H. (1999). Format dependence in subjective probability calibration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1038–1052.

Juslin, P., Winman, A., & Olssen, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard-easy effect. *Psychological Review, 107*, 384–396.

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review, 114*, 678–703.

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

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1982). Calibration of subjective probabilities: The state of the art up to 1980. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 306–334). Cambridge, UK: Cambridge University Press.

Odean, T. (1998). Volume, volatility, price and profit when all traders are above average. *Journal of Finance, 53*, 1887–1934.

Pfeifer, P. E. (1994). Are we overconfident in the belief that probability forecasters are overconfident? *Organizational Behavior and Human Decision Processes, 58*, 203–213.

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review, 33*, 7–17.

Stigler, S. M. (2002). *Statistics on the table: The history of statistical concepts and methods*. Cambridge, MA: Harvard University Press.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psyhologica, 47*, 143–148.

Svenson, O. (1978). Risks of road transportation in a psychological perspective. *Accident Analysis and Prevention, 10*, 267–280.

Svenson, O., Fischhoff, B., & MacGregor, D. (1985). Perceived driving safety and seatbelt usage. *Accident Analysis and Prevention, 17*, 119–133.

Trivers, R. (2002). *Natural selection and social theory* (Evolution and Cognition Series). New York: Oxford University Press.

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

Wason, P. C. (1966). Reasoning. In B. M. Foss (Ed.), *New horizons in psychology.* Hammondsworth, England: Penguin.

Winman, A., Hansson, P., & Juslin, P. (2004). Subjective probability intervals: How to reduce "overconfidence" by interval evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1167–1175.