Discussion

# Content-blind norms, no norms, or good norms? A reply to Vranas

## Gerd Gigerenzer*

*Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany*

## Abstract

In the psychology of thinking, little thought is given to what constitutes good thinking. Instead, normative solutions to problems have been accepted at face value, thereby determining what counts as a reasoning fallacy. I applaud Vranas (Cognition 76 (2000) 179) for thinking seriously about norms. I do, however, disagree with his attempt to provide post hoc justifications for supposed reasoning fallacies in terms of 'content-neutral' norms. Norms need to be *constructed* for a specific situation, not *imposed* upon it in a content-blind way. The reason is that content-blind norms disregard relevant structural properties of the given situation, including polysemy, reference classes, and sampling. I also show that content-blind norms can, unwittingly, lead to double standards: the norm in one problem is the fallacy in the next. The alternative to content-blind norms is not *no* norms, but rather carefully designed norms. © 2001 Elsevier Science B.V. All rights reserved.

*Keywords*: Reasoning; Probability; Norms; Rationality; Fallacy; Error

## 1. Introduction

A psychiatrist I know prescribes Prozac to his depressive patients. As do many drugs, Prozac has side effects. My friend used to inform each patient that he or she had a 30–50% probability of developing a sexual problem such as the loss of libido or peripheral malfunctioning. Upon hearing this, patients became concerned and anxious. After learning of our research on how to help people understand uncertainties, the psychiatrist changed his method of communicating risks to his patients. He

now told them that out of every ten patients to whom he prescribes Prozac, three to five experience a sexual problem. Mathematically, these two ways of communicating risks seem to be identical; psychologically, however, they are very different. Patients informed in terms of frequencies rather than probabilities were less anxious about taking Prozac. Only then did the psychiatrist realize that he had never asked what it was the patients understood by a '30–50% probability of a sexual problem'. It turned out that many of them had thought that something would go wrong in 30–50% of *their* sexual encounters. For years, my friend had simply not noticed that what he intended to communicate was not what his patients understood.

The original statement ('you have a 30–50% probability of developing a sexual problem') specifies a *single-event probability*, expressed as a percentage. By definition, however, single-event probabilities leave the *reference class* unclear. Does the percentage refer to a class of people (e.g. 'my patients who take Prozac'), to a class of events (e.g. 'my sexual encounters'), or to some other class? The psychiatrist thought of *his* patients, whereas his patients thought of *their* sexual encounters – everyone from his or her own point of view. In contrast, frequencies such as 'three out of ten patients' specify the reference class, thus reducing potential miscommunication.

## 2. Content-blind norms

Vranas (2000) says that he sees "no obvious way in which a subjectivist concept of probability is problematic". The subjectivist concept allows probabilities of single events. The Prozac example illustrates *one* problem with statements about the probability of single events: a reference class is not specified. People who think in terms of concrete cases will fill a class in, consciously or unconsciously, from their own perspective. The example also indicates that when reference classes differ, single-event probabilities and frequencies may be systematically different. To construct a norm requires analyzing the situation first, including how people construct reference classes when making judgments. In the psychology of thinking, judgment, and decision making, however, norms are rarely constructed for the problem at hand. Rather, they tend to be imposed without even looking at the content of the reasoning problem. I have called these *content-blind norms* (Gigerenzer, 1996).

The existence and nature of a unique 'correct' or 'normative' answer to a reasoning problem has often been assumed in heuristics-and-biases research uncritically. With this assessment and the need to think about norms, Vranas and I agree. Nevertheless, Vranas tries, post hoc, to rescue the 'normative' answers from my critique. His defense is that norms are, in his words, "*content-neutral*: they apply to every problem, and thus also to the Linda problem". (I will get to the Linda problem soon.) He uses the term content-neutral as I use the term content-blind. His examples of general, neutral norms include the conjunction rule, transitivity, and maximizing expected utility. In my view, however, a rule is not a norm per se, that is, it does not apply to every problem. A norm relates to a specific class of situations, and it needs to be constructed and justified for these situations.

Content-blind norms are of legitimate use within a formal system, such as for defining subjective probabilities in terms of certain rules like additivity. However, when we go beyond a formal system and want to find the best judgment or choice concerning a situation in the real world of human affairs, we have to construct norms for *this* situation, taking its characteristic structure and goals into account. In the real world, including the small world of textbook problems, the normative response depends on what we know or assume about the situation. The use of content-blind norms, in contrast, assumes that one does not have to take the situation into account.

## 3. What's wrong with content-blind norms?

Vranas asked me to explain what is wrong with content-blind justifications of probabilistic norms. I am glad to do this. With no aspiration to an exhaustive treatment, I discuss four classes of problems: polysemy, reference classes, sampling, and other structural assumptions about the situation.

### 3.1. Polysemy: not all probabilities are mathematical probabilities

The most elementary point is that words in everyday English are not the same as terms in formal logic and probability theory. For instance, the English words 'probable', 'and', 'or', and 'if … then' are not the same as mathematical probability, the logical AND, the logical OR, and the material conditional. Nevertheless, several alleged cognitive illusions are based on this confusion. In these cases, a reasoning problem is constructed as follows: (a) start with a rule of formal logic or probability; (b) replace the logical terms with terms in everyday English; and (c) add content but consider it of no normative relevance. Then it is assumed that rational reasoning consists of simply reversing this sequence. That is, the participants should not pay attention to the content of the problem, only to its everyday English terms such as 'and' and 'probable', and replace these with logical terms to derive the 'correct' answer. With that kind of rational reasoning, a person would not understand the difference between the two statements: 'Jack and Joan got married *and* they had a baby' and 'Jack and Joan had a baby *and* they got married'. By interpreting the 'and' as a logical AND, the two sentences appear to have the same meaning. However, the term 'and' is polysemous, that is, it has *several legitimate meanings*: it can mean a temporal sequence, a causal connection, or even something else. Humans can quickly infer from the content of the sentence or from the social situation which of several meanings is meant, and this content-sensitive inference reflects a form of intelligence that no computer program has mastered so far. First-order logic, in contrast, is quite simple.

Is polysemy a problem in cognitive illusions? Consider the Linda problem. Participants are given a description of a person, Linda. The description is constructed to make Linda appear to be a feminist (but not a bank teller). After reading the description, participants are asked whether it is *more probable* that Linda is 'a

bank teller' or 'a bank teller *and* is active in the feminist movement'. What, if any, is the normative answer to this problem?

Vranas, following Tversky and Kahneman (1983), subscribes to the ideal of a content-blind norm. In this view, a rational person should pay attention to only two words, the English terms 'probable' and 'and', and forget the content, including the description of Linda. Then she should replace these everyday terms with logical terms, here, mathematical probability and AND, and come to the conclusion that a conjunction of two propositions cannot be more probable than one of the propositions. She should, in this scenario, answer that it is more likely that Linda is a bank teller. Here we see a content-blind norm in action.

Is there a problem? Yes, polysemy. If one looks in the Oxford English Dictionary, or similar dictionaries, one finds that the natural language term 'probable' has multiple meanings (polysemy). Only a few of these are applicable in mathematical probability, whereas others relate to non-mathematical meanings such as plausibility, credibility, typicality, and the existence of evidence. The cover story of the Linda problem invites one to assume non-mathematical meanings of 'probable', since otherwise the description of Linda would be pointless. A person with social intelligence might infer that the experimenter would not have made her read this description if it were of no relevance. Grice's theory of conversational maxims, such as relevance, describes these social inference processes that can decode polysemy (Hilton, 1995). Here, content-blind norms are confronted with social intelligence.

This polysemy argument is testable. First, one can eliminate the polysemy of the term 'probable' by asking participants a frequency question: 'Think of 100 women like Linda. How many are bank tellers? How many are bank tellers and are active in the feminist movement?'[1] This elimination of polysemy makes the 'conjunction fallacy' largely disappear – from some 80% with probability judgments to 0–20% with frequency judgments (Fiedler, 1988; Hertwig & Gigerenzer, 1999). Another way to find out how people understand the word 'probable' is to ask them to paraphrase the Linda task to another person. Those instructed to make probability judgments used predominantly non-mathematical paraphrasings, such as 'plausible' and 'conceivable', whereas those asked to estimate frequencies used predominantly mathematical terms (Hertwig & Gigerenzer, 1999).

To summarize, English language terms such as 'probable' and 'and' are polysemous, they *legitimately* have several meanings, and only one of these has to do with mathematical probability. A content-blind norm that imposes one and only one meaning on a phrase, independent of content, is a poor guideline for intelligent behavior. Conversational norms do not conflict with statistical norms, in spite of the way Vranas sees the situation; rather they precede them, providing clues for deciding whether or not the problem is about statistical reasoning in the first place.

The 'conjunction fallacy' in the Linda problem is not a reasoning fallacy. People do well understand the relation between a set and a subset. This has been known for a long time, but psychology has a short memory. Inhelder and Piaget (1969) studied

---

[1] Note that a frequency question reduces the polysemy of the term 'probable' but not that of the term 'and'.

conjunction problems. They showed children a box containing wooden beads; most of them were brown, but a few were white. Then they asked the children 'Are there more wooden beads or more brown beads in this box?' By the age of 8 years, a majority of the children responded that there were more wooden beads, indicating that they understood conjunctions (class inclusions). Note that Inhelder and Piaget asked children for judgments of frequency ('Are there more…').

## 3.2. Reference classes

Now I will turn to situations in which polysemy is not a problem, that is, where people realize that their task is to judge mathematical probabilities. The Prozac story illustrated that single-event probabilities can be interpreted by different people in different ways, because, by definition, no reference class is specified. I will now show how this can lead to systematic differences between judgments of probabilities and relative frequencies.

Following the revival of subjective probabilities in the second half of the 20th century, probation officers have been asked to estimate the probability that an offender, say Mr Jones, will commit another violent act if given parole. This is a single-event judgment. Slovic, Monahan, and MacGregor (2000) gave members of the American Academy of Psychiatry and Law actual cases of violent patients described in one-page abstracts of Discharge Summaries. One group was asked to judge the probability that Mr Jones will commit a violent act, and the other group to judge how many, among 100 people like Mr Jones, will commit a violent act. The probability judgments were systematically higher than the frequency judgments. For instance, if the probability of a violent act was around 0.30, the relative frequency was around 20 out of 100. The same effect has been demonstrated experimentally by Gigerenzer, Hoffrage, and Kleinbölting (1991).

As in the Prozac case, a discrepancy between single-event probabilities and frequencies need not be equated with a reasoning fallacy, because single-event probabilities can be interpreted in different ways. For instance, one way to construct a reference class is: 'If Mr Jones is on weekend release 100 times, how often would he commit a crime?' Or one can think of a class of men: 'If there are 100 men like Mr Jones out on one weekend, how many of them would commit a crime?' The resulting probabilities need not be the same.[2]

To summarize, the Prozac and probation examples illustrate one reason why statements about single events and frequencies can systematically differ: single-event judgments do not specify, by definition, a reference class. PMM theory (Gigerenzer et al., 1991) models how probability and frequency tasks elicit different reference classes, which are part of different probabilistic mental models. PMM theory predicts when probabilities should or should not match frequencies, and this has consequences for when we should or should not expect calibration or under/overconfidence. One can see that psychology can actually be relevant to the

---

[2] Slovic et al. (2000) try to explain this discrepancy by the "frightening images evoked by the frequency format" (p. 290), that is, by images of harmful attacks.

construction of norms. Reference classes are a second reason why norms need to be constructed for specific situations rather than imposed in a content-blind way.

### 3.3. Sampling

The concept of a reference class helps illustrate my third point: the sampling of stimuli from a class is crucial to normative claims. For instance, a person who is asked whether New York or Rome is further south might use temperature as a cue to make an inference. Since Rome has a higher average temperature, she might guess that Rome is also further south. But temperature is not a perfect cue, and Rome and New York were selected because they are among the few metropolises for which the cue leads to a wrong prediction. When experimenters sample selectively for such pairs, participants will make a disproportionately large number of mistakes, resulting in what has been called the 'overconfidence bias'. Again, zero-overconfidence or perfect calibration cannot be imposed as *the* norm in all situations, independent of the kind of sampling used. Is selective sampling a problem for cognitive illusions?

Vranas (2000) assumes not: "experimenters are typically careful to (e.g.) stipulate random sampling". This is news to me. Consider research on the overconfidence bias in general knowledge. Before we (Gigerenzer et al., 1991) drew attention to the sampling process, there was, to the best of my knowledge, no single study that used random sampling from a defined class of problems. Rather, experimenters just somehow selected the questions. Moreover, even after we showed that PMM theory can predict how selective versus random sampling can make overconfidence disappear, appear, or even invert (depending on a second variable: single-event versus frequency judgments), many of my colleagues went on for years claiming that sampling would play no role. For instance, the *Journal of Behavioral Decision Making*, 1997, Vol. 10, No. 3, and *Organizational Behavior and Human Decision Processes*, 1996, Vol. 65, No. 5, devoted entire issues to calibration and overconfidence, where the commentaries persisted in claiming that random sampling would not affect overconfidence bias. However, Juslin, Winman, and Olssen (2000) recently analyzed 130 studies with and without random sampling to see what the evidence really is. They showed that overconfidence bias indeed disappears across all 35 studies with random sampling and that this result cannot be explained away by a hard–easy effect, as had been claimed.

The disappearance of a cognitive illusion that was and still is presented as a robust fact also speaks to another point. What Vranas calls the "empirical disagreement" is larger than it at first appears. Vranas, following Kahneman and Tversky (1996), tries to play down its size. However, overconfidence bias is not the only 'cognitive illusion' that depends on the method of sampling used (Fiedler, 1996; Gigerenzer, Hell, & Blank, 1988). To summarize, sampling is a third reason for analyzing the actual situation before constructing a norm.

### 3.4. Other structural properties

I have singled out polysemy, reference classes, and sampling because these features of reasoning problems directly concern the interpretation of single-event

probabilities. However, there are other structural properties and assumptions that can be relevant to constructing a norm. I illustrate this point briefly with three Bayesian-type problems.

### 3.4.1. The cab problem

Participants have to estimate the probability that a cab involved in a hit-and-run accident was from the Blue company, given that a witness testified 'blue'. Tversky and Kahneman (1980) assumed that the cab problem has one 'correct' (or 'normative') answer, a probability of 0.41, which can be arrived at by inserting the base rate, hit rate, and false alarm rate given into Bayes' rule. The content of the problem – such as what we know about the psychology of witness testimony – was assumed to be of no relevance to deriving the normative answer. In contrast, Birnbaum (1983) analyzed the hit-and-run situation described in the cab problem. He showed that the normative answer depends on what the witness knew about the base rates of cabs, and which strategy the witness used to set his decision criterion (in the sense of signal detection theory). Depending on what assumptions one makes about these features (which are missing in the original problem), the normative answers vary between 0.2 and 0.8. Note that the witness's knowledge and strategies are crucial to the normative analysis of the cab problem. If the content of the problem is different, for example were the data not produced by a human witness but by a mechanical device, then witness strategies and shifting criteria would play no role in a normative analysis.

### 3.4.2. The engineer–lawyer problem

Participants have to estimate the probability that a person is an engineer rather than a lawyer, given a written description of that person. This problem is an instance of a second type of Bayesian problem, in which the hit rate and false alarm rate are not given, but participants need to infer these from the situation. Tversky and Kahneman (1973) again introduced a content-blind norm. In contrast, Mueser, Cowan, and Mueser (1999) applied a generalized signal detection analysis to model the way participants infer the signal-to-noise ratio from what they assume about the task. Mueser and his colleagues showed that a context-sensitive Bayesian analysis leads to normative answers that differ from the one resulting from the use of Kahneman and Tversky's content-blind norm.

### 3.4.3. The Monty Hall problem

A contestant in a game show is given the choice among three doors, behind one of which is the grand prize (a Cadillac), and behind the others are fun prizes (goats). The contestant chooses door A; then show master Monty Hall opens door B revealing a goat, at which point Monty asks the contestant whether she wants to stay with her choice or switch to door C. As in the previous examples, the correct answer – stay or switch – depends on the knowledge one has of the situation and the assumptions one has to make in the absence of full knowledge. As before, these assumptions were not specified in the original problem version. For instance, it is essential to know whether or not Monty always offers a candidate the possibility of switching

doors. If he were to offer switching only when the candidate was standing in front of the winning door, the normative strategy would be not to switch. (Monty Hall actually did not always offer the candidates the possibility of switching, and it is unclear when he chose to do so.) A Bayesian analysis can result in both recommendations – stay or switch – dependent on this and other structural features of the situation (Friedman, 1998).

## 4. Conflicting statistical norms

Each of the three major interpretations of probability – degrees of belief, propensities, and frequencies – has its characteristic virtues and vices. There is no winner. And each of the tools designed for inductive inference – Fisher's significance testing, the Neyman–Pearson hypothesis testing, and Bayesian theory, to name only a few – has its virtues and its vices. Statistical theories do not simply diverge in the third decimal place, but in the very questions they ask (Gigerenzer et al., 1989).[3]

I have pointed out that frequentism, among other things, does not support the idea that every statement about a single event must follow the laws of probability. As a defense – or as a distraction – proponents of the heuristics-and-biases program attributed tough-minded frequentism and 'normative agnosticism' to me (e.g. Kahneman & Tversky, 1996, p. 586). These attributions contradict one another and both are nonsense. Vranas (2000), who is otherwise careful, follows suit and tentatively suggests that some of my writings "may give the impression" that I would endorse the proposition that "frequentism is true". Not surprisingly, Vranas could not find any statements supporting this in my writings; what he does cite are two facts about the history and the current spread of the frequency interpretation. Similarly, despite Vranas' realization that my normative standards may be higher than those in the heuristics-and-biases program, he still seems to take Kahneman and Tversky's attribution to me of 'normative agnosticism' seriously. But the choice is not between *content-blind* norms and *no* norms. There are *good* norms, too.

## 5. Conflicting content-blind norms

An unnoticed but notable consequence of the use of content-blind norms is that they can lead to conflicting norms. That is, the norm in one problem is the fallacy in the next. How can this happen? Remember, in the heuristics-and-biases program one creates a reasoning problem by starting with a rule of logic or probability – such as regression towards the mean – and adds content, but does not analyze that content to determine the normative answer. Thus, when one starts with a different rule – say,

---

[3] A recent example of where Fisherian and Bayesian views lead to opposite conclusions is the debate about how the size of a DNA database influences the confidence of whether the person whose DNA matched the sample is actually the source of the DNA trace found on the victim. A Fisherian perspective leads to the conclusion that the larger the database through which the search was conducted, the smaller the confidence, whereas a Bayesian perspective suggests the opposite result (see Redmayne, 1998).

calibration – to create a second reasoning problem, and then adds content that, again, is not analyzed for its normative relevance, it can happen that the first norm is relevant to the second problem, contradicting this problem's content-blind norm. Within this program, however, this norm conflict is hard to recognize. Here are some examples.

### 5.1. Regression towards the mean

In situations of uncertainty (e.g. partial ignorance or unpredictability), judgments and other variables exhibit regression towards the mean. If a person does not expect events to regress – to use an example reported by Kahneman, that an inexperienced pilot's first great landing will likely be followed by a more mediocre performance – this is called a reasoning fallacy. Consider now a different part of the heuristics-and-biases program. Perfect calibration and zero-overconfidence have been imposed as norms in research on confidence in general knowledge. This knowledge, however, involves partial ignorance. Thus, an analysis of the situation would reveal that perfect calibration conflicts with another norm, regression towards the mean. Years of attributing miscalibration and overconfidence bias to people passed before it was pointed out that these so-called fallacies can be a direct consequence of regression towards the mean (e.g. Erev, Wallsten, & Budescu, 1994; Juslin et al., 2000). Thus, in one body of research judgments in accordance with regression are considered normative, whereas in another body they are considered a fallacy.

### 5.2. Sample size

When one observes that the ratio of heads to tails is 7:3, one should be more confident that the coin is actually biased when the sample size is 100 rather than 10. Disregarding sample size has been called a reasoning fallacy in one group of studies on what has been labeled the law of small numbers (see Sedlmeier & Gigerenzer, 1997). Compare this with a different body of research, which starts with Bayes' rule as a norm. Note that the simple form of Bayes' rule used in the heuristics-and-biases research disregards the advantages of beta distributions, which reflect the reliability of base rates and likelihoods. A major determinant for this reliability is sample size. For instance, the sample size from which the base rate is calculated makes no difference to the normative computations using Bayes' rule in the heuristics-and-biases program. Thus, in one body of research attention to sample size is called normative, whereas in the other inattention to sample size is part of the norm.

### 5.3. Hypotheses testing

In one body of research on inductive inference from data to hypotheses, Bayes' rule is said to be the norm. In a related body of research (inspired by Neyman–Pearson theory), the norm is that data should always be compared to at least two hypotheses, not just to one (e.g. McKenzie, 1998). When participants do not follow Bayes' rule, a reasoning fallacy is diagnosed, such as the base rate fallacy. When participants do not consider an alternative hypothesis, another reasoning fallacy is

diagnosed, such as confirmation bias. Now compare these two norms with how researchers themselves make inductive inferences from data to hypothesis, such as when testing the hypothesis that participants reason according to Bayes' rule or pay attention to alternative hypotheses. Researchers routinely use significance testing, not Bayes' rule or Neyman–Pearson hypotheses testing. In significance testing, one always ignores base rates and computes the likelihood of the data against only one hypothesis (the null hypothesis). Thus, many researchers' norm of hypotheses testing contains the fallacies they diagnose in their experimental participants.

You may call this a double standard. Most striking to me, however, is that most of my colleagues to whom I pointed this out were not aware of their double standard. And after they did become aware of it, most did not change their behavior. Norms, one said, tell you what you are required to do; you don't think about them.[4]

There is a positive lesson to be learned from these examples. We need to start teaching the next generation of researchers how to think about norms. Normative thinking involves the exercise of thought, not the imposition of content-blind norms. It can have the charm and challenge of detective work: ascertaining the facts of a situation, making assumptions about what one does not know, and then deriving the consequences.

## 6. Afterthought

Let me end on a more general note. The problems Vranas and I have debated are word problems. Their solutions are based on the assumption that all relevant information is already on the table and is completely believable and reliable. Norms in the real world, however, need to be designed for complex situations where information must be searched for and is often unreliable, where time is money, and where deadlines approach rapidly. Here, sound normative thinking leads us into the world of bounded rationality, of fast and frugal heuristics, satisficing, and other robust strategies that can do surprisingly well when used in the appropriate situation (Gigerenzer, 2000; Gigerenzer, Todd, & the ABC Research Group, 1999). These strategies do not follow the classical norms, such as maximization, but rather psychological principles like recognition and limited search. This work shows, even more than the textbook problems we have discussed, that psychology is indispensable to good thinking.

## References

Birnbaum, M. H. (1983). Base rates in Bayesian inference: signal detection analysis of the cab problem. *American Journal of Psychology*, *96*, 85–94.

---

[4] This list of conflicting norms can be extended to research on deductive thinking. For instance, many researchers studying the Wason selection task have taken for granted that the laws of the material conditional would specify the normative answer to the selection task, independent of its content. However, Oaksford and Chater (1994) showed that a Bayesian analysis can favor answers as normative that have been called fallacies for decades.

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.

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

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*, 193–214.

Friedman, D. (1998). Monty Hall's three doors: construction and deconstruction of a choice anomaly. *The American Economic Review*, *88* (4), 933–946.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky (1996). *Psychological Review*, *103* (3), 592–596.

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

Gigerenzer, G., Hell, W., & Blank, H. (1988). Presentation and content: the use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, *14* (3), 513–525.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: a Brunswikian theory of confidence. *Psychological Review*, *98* (4), 506–528.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: how probability changed science and everyday life*. Cambridge: Cambridge University Press.

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

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: how intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, *12* (4), 275–305.

Hilton, D. J. (1995). The social context of reasoning: conversational inference and rational judgment. *Psychological Bulletin*, *118*, 248–271.

Inhelder, B., & Piaget, J. (1969). *The early growth of logic in the child*. New York: Norton Library.

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* (2), 384–396.

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

McKenzie, C. R. M. (1998). Taking into account the strength of an alternative hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 771–792.

Mueser, P. R., Cowan, N., & Mueser, K. T. (1999). A generalized signal detection model to predict rational variation in base rate use. *Cognition*, *69*, 267–312.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101* (4), 608–631.

Redmayne, M. (1998). The DNA database: civil liberty and evidentiary issues. *Criminal Law Review*, *July*, 437–454.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: the empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*, 33–51.

Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: the effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior*, *24* (3), 271–296.

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

Tversky, A., & Kahneman, D. (1980). Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1, pp. 49–72). Hillsdale, NJ: Erlbaum.

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

Vranas, P. B. M. (2000). Gigerenzer's normative critique of Kahneman and Tversky. *Cognition*, *76*, 179–193.