The Handbook of Rationality

Edited by: Markus Knauff, Wolfgang Spohn

Citation:

The Handbook of Rationality
Edited by: Markus Knauff, Wolfgang Spohn
DOI: 10.7551/mitpress/11252.001.0001
ISBN (electronic): 9780262366175
Publisher: The MIT Press
Published: 2021

Funding for the open access edition was provided by the MIT Libraries Open Monograph Fund.

15.3 How to Improve Rational Thinking?

Stephanie de Oliveira and Richard Nisbett

Summary

Current perspectives on what constitutes "good thinking" suggest that it should include proper process (e.g., be internally coherent/rational) and some pragmatic element of usefulness (e.g., it should produce actions and results that are accurate and/or beneficial to the thinker). Although people's reasoning in economic and social domains has various shortcomings on both counts, it can be improved through training. Such training does not need to be intensive or time-consuming in order to produce measurable improvements in reasoning. However, it typically must go beyond simply telling people to be more rational, to "try harder," or to avoid particular biases. The most effective training seems to be the type that uses diverse examples and offers practice and feedback opportunities. Although some research demonstrates correlational benefits of "good thinking" for various life outcomes, future work should examine whether training has causal effects on improving outcomes for people over time.

1. What Is Good Thinking?

Consider the following scenario:

> Imagine you spent $100 on a ticket for a ski trip to Michigan. Several weeks later you buy a $50 ticket for a ski trip to Wisconsin. You think you will enjoy the Wisconsin ski trip more than the Michigan ski trip. As you are storing your Wisconsin ski trip ticket in your wallet, you notice that both trips are for the same weekend. It's too late to sell either ticket, and you cannot return either one. You must use one ticket and not the other. Which trip will you take? (adapted from Arkes & Blumer, 1985, p. 126).

Which is the better decision? Why?

Now, consider the following statements:

> If it's raining, then the streets must be wet.
> It's not raining.
> Therefore, the streets must not be wet. (Nisbett, 2015, p. 218)

Is it better to agree or disagree with the conclusion? Why?

Such exercises help us grapple with questions about what constitutes normative thinking. Are there better and worse ways of thinking? And if there are better ways of thinking, how can people improve? We begin this chapter with some definitions of good reasoning that have been popular across different cultures and disciplines. We then summarize research on human judgment in the social sciences, focusing on diagnosis ("How good is human judgment?") and potential solutions ("Can it be improved?").

Greek philosophers played a major role in defining "good thinking" in Western societies. Their rules of logic are used to this day. The second opening example of this chapter shows the commonly made "inverse error"—reasoning that violates the rules of propositional logic. The first argument takes the form "If *P* then *Q*." This means that if we observe *P* (rainy weather), we can conclude *Q* (wet streets). The second statement says that we observe "*not P*." Firm conclusions cannot in fact be drawn because *P* did not occur.

Conclusions can be logically derived yet wrong, and they can be illogically derived yet right. The internal validity of an argument and its truth are two entirely separate things. Thus, modern psychologists' definitions of good reasoning typically take both factors into account. One prominent perspective states that good reasoning should include "correspondence" and "coherence" (Hammond, 1996); its conclusions should generally *correspond* with reality, and it should also use *coherent* processes to produce such conclusions. As another prominent psychologist put it, "Good judges should both 'get it right' and 'think the right way'" (Tetlock, 2005, p. 7).

Nevertheless, people in different societies have very different ideas about what constitutes proper, or normative, reasoning. For example, dialectical systems of thought tend to emphasize the shifting nature of reality, acknowledging that reality is full of contradictions, changes, and coexisting yet opposing forces (DaMatta, 1995; Nisbett, Peng, Choi, & Norenzayan, 2001; Peng & Nisbett, 1999). Correspondence and coherence may be hard or even impossible to achieve, since what is true at one moment may be false a moment later. The dialectic system of thought is more heavily adopted in some East Asian and Latin American societies (de Oliveira & Nisbett, 2017a; Nisbett et al., 2001), and the difference may be partially traceable to differences in how philosophy developed. For example, in contrast to Greek philosophy, which was developed in the service of guiding debate, East Asian philosophy was developed to serve society. Thus, good thinking was not necessarily about being right in some logical abstract sense but about managing the vicissitudes of life in a pragmatically useful and socially harmonious way.

Some Westerners have also argued that thinking should not be assessed merely by its adherence to abstract principles (Keys & Schwartz, 2007). Instead, one should consider how decision outcomes are experienced by the thinker and "how the decision-making process fits into the decision maker's life as a whole" (p. 165). From an economic perspective, the correct answer to this chapter's opening question about the ski tickets is to go on the trip that one would typically enjoy more since the cost of the $100 ticket cannot be recovered. Going on the less enjoyable, more expensive trip amounts to wasting time in addition to wasting money since one has thereby forgone a more enjoyable way to spend one's vacation. Yet most untrained people who answer that question choose the more expensive and less enjoyable trip to Michigan. Although that decision does not adhere to economic principles, choosing Wisconsin may make people feel like they are violating another cherished rule, "Waste not, want not" (Arkes & Ayton, 1999). The unpleasantness of violating that rule may carry over from the decision process into the decision experience, making Wisconsin the less appealing choice at the moment even if it is the more appealing choice when detached from the decision dilemma (Keys & Schwartz, 2007).

To summarize, there are diverse perspectives on what constitutes good thinking across cultures and within societies. Most psychological research on judgment and decision making has studied Western thinking and evaluated it by Western standards of coherence and correspondence. Thus, our chapter adopts a similar Western focus, but we have reviewed important cultural variation in reasoning elsewhere (de Oliveira & Nisbett, 2017b; Yates & de Oliveira, 2016).

2. How Good Is Human Judgment?

Psychologists began to empirically study reasoning relatively recently. Initial studies from the 1960s through the 1980s suggested that although human judgment has been adequate for survival, it is characterized by errors in logic and nonadherence to normative economic principles. In social judgment studies, people paid too much attention to the actor and not enough attention to the actor's context (Ross, 1977). They based social judgments and frequency estimates not on base rates or valid predictors but on whether examples of similar cases easily came to mind (Tversky & Kahneman, 1973). In economic studies, people were loss-averse, leading them to make contradictory decisions when identical options were framed either as gains or as losses (Kahneman & Tversky, 1982). People were also poor logicians. For example, they were terrible at testing conditional propositions. Wason's card selection task tested how well people could falsify *if P then Q* propositions (Wason, 1960). In the task, people received a rule (e.g., "If the front of the card has a vowel, then the back has an even number"). Participants would then examine four cards displaying letters and numbers (e.g., an A, a B, a 4, and a 7). Their task was to turn over only as many cards as necessary to test whether the rule was violated. Few people would correctly turn over the A card and the 7 card.

Compendiums of reasoning errors were published in books like *Human Inference* (Nisbett & Ross, 1980). Such works were not optimistic about human reasoning: "We insist that the errors demonstrated in the laboratory and chronicled in this book are the ingredients of individual and collective human tragedy" (p. 251). But after several years of experiments demonstrating reasoning errors, Nisbett and his colleagues failed to replicate one of Kahneman and Tversky's findings. Specifically, for a problem in which people typically misapplied the law of large numbers, they found that a large share of their participants responded accurately (Nisbett, 1993, p. 4). Participants who had taken a statistics course did particularly well, whereas those who had not taken statistics performed poorly, like Kahneman and Tversky's original participants.

Taking a different approach, another scholar found that although people were bad at the standard Wason card selection task, they were better at testing conditional propositions when the task mapped onto certain real-world scenarios. For example, when asked to turn over as many sales receipts as necessary to establish that, if the receipt is for more than $20, it has a signature on the back, people largely understood that they should check receipts with large amounts and unsigned backs (D'Andrade, 1982, as cited in Lehman, Lempert, & Nisbett, 1988).

These findings indicated that mental tools necessary to engage in sound reasoning were not altogether *absent* from the human toolbox. Rather, they were neglected for certain problems, and absent concepts could perhaps be taught via formal training. Scholars hoping to improve cognition would need to be able to teach new strategies but also to find ways of encouraging the use of already existing strategies and rules across domains.

In summary, there is ample research documenting reasoning "failures" in humans, but there are diverse views on just how much of a problem this is and on whether observations of such failures are partly due to how the questions are asked. Taking abstract logical tasks like Wason's card selection task and couching them in more practical terms can improve people's performance. Nevertheless, people—even experts—still make many errors in common everyday judgments. Judgment errors can materialize in social domains, in health care (e.g., Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007), in economic domains (e.g., Arkes & Blumer, 1985), and in political domains (Tetlock, 2005). Although we are less gloomy about the state of reasoning than psychologists were in the 1980s, there is substantial room for improvement.

3. Can Reasoning Be Improved?

There have been many proposed approaches to improving reasoning (for a general review, see Larrick, 2004). Approaches include changing the structure of the question or the task (e.g., Thaler, Sunstein, & Balz, 2012), using groups rather than individuals (Sunstein & Hastie, 2015; Surowiecki, 2005), or even jettisoning the human altogether in favor of mathematical models (Dawes, 1979). In this chapter, we focus on training.

Consider the following problem (adapted from Gigerenzer et al., 2007):

> Assume you conduct breast cancer screening in a certain region. You know the following information about this region: (1) The probability that a woman has breast cancer is 1%; (2) if a woman has breast cancer, the probability that she tests positive is 90%; and (3) if a woman does not have breast cancer, the probability that she nevertheless tests positive is 9%. Now imagine that a woman tests positive. What are the chances that she has breast cancer?

Laypeople and professionals struggle with this type of problem. In a survey of 160 gynecologists, only 21% responded correctly when given four answer choices (Gigerenzer et al., 2007). (She has a 10% chance; out of 10 women with a positive mammogram, about 1 has cancer.) Those doctors received a training session in which they were taught to convert percentages into frequencies (e.g., think of 10% as 1 in 10, or 10 out of 100). Afterward, 87% were able to give the correct answer.

Research on training has generally found that reasoning can be improved through training and that this training does not need to be terribly extensive. We review some representative studies below that also address additional questions: Are some types of training more effective than others? How long do effects last? Can training on one rule in one domain transfer such that those newly learned skills are more broadly applied to other problems? We begin with some findings from statistical reasoning and then report some results on how the environment and feedback can improve rational reasoning.

3.1 Learning Statistics

Although early psychologists were skeptical about the effectiveness of training, research on training has produced encouraging results. For example, one study assessed people's reasoning by asking them to explain why a traveling saleswoman's second visit to a restaurant typically leads to disappointment when her first meal had been outstanding (Fong, Krantz, & Nisbett, 1986). On such questions, respondents with no statistical background gave exclusively nonstatistical answers (e.g., suggesting that the chefs change a lot or that her expectations were too high). Respondents who had taken one statistics course gave statistical answers about 20% of the time (e.g., suggesting she was lucky the first time, noting that restaurant meal quality varies). New graduate students in psychology gave statistical answers about 40% of the time, and doctoral-level scientists at a research institution gave such answers about 80% of the time. Giving a statistical answer to questions such as these reflects better thinking because single extreme observations are typically followed by less extreme observations; extreme events "regress" toward the mean value more often than not.

Further research suggested that the effects of formal training on reasoning depended on how much the program linked abstract principles, including statistical ones, to everyday life problems. Skills taught in a domain-restricted manner may show less generalizability than skills taught with examples of diverse applications. Lehman and colleagues (1988) found that getting a graduate degree in medicine or psychology was associated with substantial increases in conditional reasoning and statistical/methodological reasoning, but getting a graduate chemistry degree was not. Chemistry students typically study very little statistics or scientific methodology with direct relevance to everyday life. Psychology and medical students study a great deal of statistics and scientific methodology having relevance to everyday-life domains.

This idea—that people can generalize learned reasoning rules across domains if they are shown how and get practice examples with it—was tested in a more formal experiment (Fong et al., 1986). In their first study, the researchers tested people's reasoning through problems that varied from seeming relatively objective (e.g., judging the characteristics of a lottery) to relatively subjective (e.g., choosing which college to attend). Participants were randomly assigned to one of five conditions. In one condition, people read a packet explaining statistical principles such as the law of large numbers. (The law of large numbers in statistics states that smaller samples are more likely than larger samples to vary from the population distribution.) After reading the packet, they watched the researcher demonstrating the principle by drawing balls from an urn. In another condition, participants worked through example problems in which those abstract principles were applied to everyday problems. For example, they read about hypothetical ballet auditions and learned to think of the observed dance moves as a sample of a larger "population" of dance moves. They were reminded that smaller samples—like dance moves observed at an audition—are unreliable estimates of population distributions—such as all of a dancer's moves. They then considered how, based on this observation, it would make sense that some of the best dancers at an audition would turn out to be average performers after all; their good performance at the audition was a statistical fluke due to small sample size. In a third condition, participants read the informational packet *and* worked through the example problems. There were two control conditions; in one, participants were given no training. In another, they were simply given the definitions of each of the statistical principles.

Training influenced the frequency and quality of statistical responses. In both control conditions, fewer than half of the responses invoked statistical principles, and the quality of the statistical answers they gave was low. In the full training condition, about 64% of responses invoked statistical principles, and the quality of the answers was relatively high. The examples-only training and the packet-only training conditions fell in between.

A similar study tested whether training in one domain generalized to other domains and whether effects persisted over time (Fong & Nisbett, 1991). Participants were trained on the law of large numbers through example problems. They were tested immediately or after a two-week delay, either in the same domain as the examples or in a different domain. The training was equally effective immediately whether participants were tested in the same or a different domain. Participants tested later in the same domain showed no decline in their performance. Participants tested later in a different domain showed some decline, but they still performed better than participants who had received no training. Similar work examining financial training found domain-independent gains as well: people trained in normative economic reasoning with financial (vs. nonfinancial) examples did better on both financial and nonfinancial measures of reasoning (Larrick, Morgan, & Nisbett, 1990), and training gains were still observed four to six weeks post-training. It's notable that both statistical training and cost–benefit analysis training improved answers to questions in non-laboratory, nonacademic settings such as alleged telephone surveys about politics or consumer preferences.

3.2 The Learning Environment and Feedback

Competence is best developed in contexts where the learners can get relatively fast, frequent, and complete feedback on their performance—this allows people to know whether there is a problem with their reasoning in the first place (Hogarth, 2001; Larrick, 2004). For example, if one wants to know how good one is at spotting academic talent in the graduate admissions process, one would need to track outcomes of one's decisions—both the accepted and rejected students—over many admissions cycles. Some real-world tasks provide people with feedback-heavy learning environments (e.g., weather forecasting, sports training), but many tasks provide slow and incomplete feedback (e.g., admissions and hiring decisions).

Training programs can artificially structure the learning environment so that it provides fast and complete feedback on tasks that otherwise are learned in feedback-poor real-world contexts. Morewedge and colleagues (2015) compared the effectiveness of video training and video *game* training for improving reasoning. The video game allowed participants to not only learn about the biases but also engage in problem-solving tasks that elicited those biases. After responding to the task, they received feedback on their performance as well as suggestions for how to improve. Participants trained with videos merely received information about those biases and did not receive performance feedback. The interventions targeted bias blind spot (i.e., thinking you are less biased than others), confirmation bias (i.e., overweighting confirming vs. disconfirming evidence), the fundamental attribution error (i.e., neglecting situational influences when attributing events to a person), anchoring (i.e., overweighting initial information when making a subsequent judgment), the representativeness heuristic (i.e., relying on judgments of similarity to make inferences about probability), and social projection (i.e., assuming others' emotions and values are like one's own).

In both studies, comparisons of pre- and post-test reasoning revealed that both the video and the video game reduced bias. Follow-up assessments two to three months later showed that the improvements were sustained for both methods. However, the video game debiasing effects were much larger than the video effects at both post-intervention assessment times. These results suggest that personal feedback and practice greatly enhance the effects of training.

One research team recently demonstrated increases in geopolitical forecasting competence both by training people with instructional modules and also by placing workers in more interactive, feedback-rich team environments (Mellers et al., 2014). The training participants received either 45-minute "scenario training" or "probability training" of similar length. The "scenario training" module taught forecasters to generate new futures, actively entertain more possibilities, use decision trees, and avoid certain biases. The "probability training" module taught forecasters to average multiple estimates from models, polls, and expert panels; extrapolate over time when variables were continuous; and avoid errors like base-rate neglect. Forecasters working in teams (vs. individually) had ample experiential learning opportunities; they could make forecasts as often as they wished, interact with other forecasters, share information via a web-based platform, get feedback on their thoughts and on their forecasts, and learn from that feedback. Training improved performance in both the first year and the second year of the tournament. Teaming also helped: forecasters working in teams did better than those working individually. Remarkably, although the authors did not offer formal statistical analyses on the matter, the benefit of teaming seemed to be larger than that of training.

4. Conclusion

Training can improve reasoning, at least by Western definitions, and the type of training matters. Training with examples and practice helps people generalize specific skills to a variety of domains, and feedback-rich environments likely boost training effects and help maintain gains over time. Programs do not need to be intensive or long in order to produce lasting effects on reasoning, but they do need to go beyond merely telling people about common judgment errors or instructing them in abstract principles.

There is good reason to believe training effects would carry over to everyday life problems that do not closely resemble problems used for training. Research has linked higher normative reasoning ability to lower risky behavior and juvenile delinquency (Parker, Bruine de Bruin, Fischhoff, & Weller, 2018) as well as higher salaries for a sample of professors and higher GPA for a sample of students (Larrick, Nisbett, & Morgan, 1993).

References

Arkes, H. R., & Ayton, P. (1999). The sunk cost and Concorde effects: Are humans less rational than lower animals? *Psychological Bulletin*, *125*(5), 591–600.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*, 124–140.

DaMatta, R. A. (1995). For an anthropology of the Brazilian tradition, or "A virtude está no meio." In D. J. Hess & R. A. DaMatta (Eds.), *The Brazilian puzzle: Culture on the borderlands of the Western world* (pp. 270–292). New York, NY: Columbia University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582.

de Oliveira, S., & Nisbett, R. E. (2017a). Beyond East and West: Cognitive style in Latin America. *Journal of Cross-Cultural Psychology*, *48*(10), 1554–1577.

de Oliveira, S., & Nisbett, R. E. (2017b). Culture changes how we think about thinking: From "Human Inference" to "Geography of Thought." *Perspectives on Psychological Science*, *12*(5), 782–790.

Fong, G. T., Krantz, D. H., & Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, *18*(3), 253–292.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General*, *120*(1), 34–45.

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

Hammond, K. R. (1996). *Human judgment and social policy: Irreducible uncertainty, inevitable error, unavoidable injustice*. New York, NY: Oxford University Press.

Hogarth, R. M. (2001). *Educating intuition*. Chicago, IL: University of Chicago Press.

Kahneman, D., & Tversky, A. (1982). The psychology of preferences. *Scientific American*, *246*(1), 160–173.

Keys, D. J., & Schwartz, B. (2007). "Leaky" rationality: Normative standards of rationality. *Perspectives on Psychological Science*, *2*(2), 162–180.

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–337). Maiden, MA: Blackwell.

Larrick, R. P., Morgan, J. N., & Nisbett, R. E. (1990). Teaching the use of cost–benefit reasoning in everyday life. *Psychological Science*, *1*(6), 362–370.

Larrick, R. P., Nisbett, R. E., & Morgan, J. N. (1993). Who uses the cost–benefit rules of choice? Implications for the normative status of microeconomic theory. *Organizational Behavior and Human Decision Processes*, *56*, 331–347.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday life events. *American Psychologist*, *43*, 431–443.

Mellers, B. A., Ungar, L. H., Baron, J., Ramos, J., Gurcay, B., Fincher, K., . . . Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science*, *25*(5), 1106–1115.

Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing decisions: Improved decision making with a single training intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129–140.

Nisbett, R. E. (1993). Reasoning, abstraction, and the prejudices of 20th-century psychology. In R. E. Nisbett (Ed.), *Rules for reasoning* (pp. 1–14). Hillsdale, NJ: Erlbaum.

Nisbett, R. E. (2015). *Mindware: Tools for smart thinking*. New York, NY: Farrar, Straus and Giroux.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review*, *108*(2), 291–310.

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

Parker, A. M., Bruine de Bruin, W., Fischhoff, B., & Weller, J. (2018). Robustness of decision-making competence: Evidence from two measures and an 11-year longitudinal study. *Journal of Behavioral Decision Making*, *31*, 380–391.

Peng, K., & Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *American Psychologist*, *54*(9), 741–754.

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

Sunstein, C. R., & Hastie, R. (2015). *Wiser: Getting beyond groupthink to make groups smarter*. Boston, MA: Harvard Business Review Press.

Surowiecki, J. (2005). *The wisdom of crowds*. New York, NY: Anchor Books.

Tetlock, P. E. (2005). *Expert political judgment*. Princeton, NJ: Princeton University Press.

Thaler, R. H., Sunstein, C. R., & Balz, J. P. (2012). Choice architecture. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 428–439). Princeton, NJ: Princeton University Press.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Yates, J. F., & de Oliveira, S. (2016). Culture and decision making. *Organizational Behavior and Human Decision Processes*, *136*, 106–118.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Knauff, Markus, editor. | Spohn, Wolfgang, editor.
Title: The handbook of rationality / edited by Markus Knauff and Wolfgang Spohn.
Description: Cambridge : The MIT Press, 2021. | Includes bibliographical references and index.
Identifiers: LCCN 2020048455 | ISBN 9780262045070 (hardcover)
Subjects: LCSH: Reasoning (Psychology) | Reason. | Cognitive psychology. | Logic. | Philosophy of mind.
Classification: LCC BF442 .H36 2021 | DDC 153.4/3—dc23
LC record available at https://lccn.loc.gov/2020048455