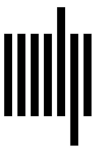
The Handbook of Rationality

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.



The MIT Press

8.6 Reasoning, Rationality, and Metacognition

Valerie A. Thompson, Shira Elqayam, and Rakefet Ackerman

Summary

In this chapter, we explore how the study of metacognition advances our understanding of human rationality. "Metacognition" refers to the processes by which we monitor and control our ongoing cognitive activities. These monitoring processes manifest as feelings on the continuum between certainty and uncertainty, which then are the stimulus for control processes (e.g., a feeling of uncertainty in an answer is a stimulus to rethink that answer, search online). Because feelings of (un)certainty are based on heuristic cues, such as the ease with which an answer comes to mind, they can mislead our control processes and produce unwarranted levels of confidence in our decisions. We also argue that metacognitive principles can help us to understand why people are not the cognitive misers that they are assumed to be and why they continue to waste cognitive resources on difficult or unsolvable problems.

1. Reasoning, Rationality, and Metacognition

Here, we explore how understanding metacognitive processes can shed light on issues of rationality. At first glance, the connection between metacognition and rationality may seem straightforward. Metacognition, after all, is typically understood to be "thinking about one's thinking," a process that seems integral to achieving rational decisions.[1] However, most researchers are less concerned with this self-reflective stance and more concerned with the ubiquitous processes by which we monitor and control our ongoing cognitions. By "monitoring," we mean the processes that keep track of how well we are doing; control processes include changing strategies, maintaining course, or giving up.

Monitoring and control again sound like high-level, reflective processes, and there is no doubt that they can be. In most ordinary instances, however, they operate passively (Koriat, 2007), in much the same way as a thermostat passively monitors the temperature of a room. The associated control process would be to send a signal to the furnace to start producing warm air when the temperature drops below a certain threshold (Thompson, 2016). Examples include the following:

1) You meet someone whose face you recognize but whose name you are not certain of. You decide not to address her by name.
2) You order a desk from a furniture store and it arrives, unassembled. Your first thought is "I can't put that together." You then look at the instructions, reconsider, and begin assembling. Halfway through, you realize that your first judgment was right and call a friend to help.
3) You are working on writing a manuscript and are dissatisfied with the progress. Even though the words accurately convey your thoughts, you think that it does not have that certain ring that will convince readers about the points you wish to make. You decide to take a break and make a fresh start in the morning.

In each of these examples, a set of cognitive processes is unfolding, involving memory retrieval, solving a problem, writing a manuscript, and a second set of processes layered on top, which we refer to as "monitoring and control processes." The monitoring processes are assumed to be continuous and, as above, to manifest primarily as a sense of (un)certainty about how well things are proceeding. In this chapter, we will primarily be concerned with monitoring and control processes associated with reasoning, problem solving, and decision making, an area of study termed "Meta-Reasoning" (Ackerman & Thompson, 2015, 2017).

The meta-reasoning framework posits many different kinds of monitoring judgments. A *judgment of solvability* is an initial assessment of whether a problem is solvable and whether one personally is capable of solving it, as the self-assembly example in number 2 illustrates. The *feeling of rightness* (FOR) accompanies an initial, intuitive answer to a problem, as illustrated by example 1.

Examples 2 and 3 illustrate the ongoing nature of monitoring and the assessment of intermediate confidence. At the end of a process, one is left with a final sense of confidence or feeling of error that one has successfully or unsuccessfully completed the task at hand.

The reason these monitoring judgments matter is their link to control processes: feelings of certainty and uncertainty are the arbiters of action (Thompson, 2016). A feeling of certainty is a signal that all is well and to carry on with the present course of action. Uncertainty signals the need to reassess. For example, reasoning problems that cue a strong FOR to an initial answer receive less reanalysis than problems that cue a low FOR, whereas a low FOR is a cue to look for alternative answers (Thompson, Prowse Turner, & Pennycook, 2011); similarly, low intermediate confidence is a signal to give up (Ackerman, 2014) or seek additional resources, as in the self-assembly example, and initial judgments of solvability (example 2) predict persistence in problem solving (Lauterman & Ackerman, 2019).

The trouble is that these monitoring judgments are usually made implicitly, on the basis of heuristic cues (Koriat, 1997). Heuristic cues can be based on general self-perceptions (e.g., "I am no good at DIY projects"), global information about the task (e.g., "These instructions seem particularly opaque"), and characteristics of individual items (see Ackerman, 2019, for a review). One item-based cue is fluency, or the ease with which an answer comes to mind. Fluency is a cue to confidence and FOR (Ackerman & Zalmanov, 2012; Thompson, Evans, & Campbell, 2013), regardless of accuracy. That is, answers that come to mind fluently are experienced positively (Thompson & Morsanyi, 2012) and lead to a sense of certainty, as does familiarity (Reder & Ritter, 1992).

The implications for human rationality are clear: because the monitoring processes can send misleading signals, control processes may not operate in an optimal manner. That is, one might believe that a problem is solvable, leading one to persist in trying to solve it, even when it is not (Lauterman & Ackerman, 2019). Similarly, one might have a strong FOR in a suboptimal answer and thus fail to think further; conversely, one might have a weak FOR in a good solution and thus spend unnecessary time thinking it over (Thompson et al., 2011; Thompson & Johnson, 2014). Indeed, one of the reasons that people commit errors on traditional heuristics-and-biases tasks is that the first answer tends to come to mind fluently and is therefore held with high confidence (Thompson et al., 2013). As a result, one may fail to reflect further on the answer and give an answer based on beliefs, stereotypes, or familiarity, when the correct answer requires one to think logically or probabilistically.

In this chapter, we consider the advantages and drawbacks of the proper allocation of deliberate thinking and the role of normative theories, such as logic and probability, as markers of rationality. As stated earlier, the goal of this chapter is to consider how understanding metacognitive processes can advance our understanding of human rationality. In doing so, we examine how metareasoning and metacognition shed new, and sometimes surprising, light on rationality, when considered from a variety of perspectives. In addition to normative correctness, we consider issues related to pragmatic rationality, bounded rationality, and epistemic rationality.

2. Overconfidence and Epistemic Rationality

Epistemic rationality is usually understood to mean having true beliefs about the world (i.e., believing true propositions and disbelieving false ones).[2] In this section, we demonstrate how miscalibrated judgments of confidence play a role in acquiring and maintaining false beliefs about the world and the manner in which one evaluates one's knowledge of the world.

The *feeling of knowing* is one type of metacognitive judgment we make in evaluating our own knowledge. For example, one may be quite sure that one will ultimately be able to remember the capital city of Uruguay, even if it currently eludes recall. These feelings are often inferences that rely on cues such as the amount of information that comes to mind (Koriat, 1993), for example, it starts with M, is on the coast, and so on. Dunning (2011) argued that these cues underlie a set of phenomena he termed "meta-ignorance": an ignorance about the things one does not know about. Examples of meta-ignorance include *overclaiming*, in which people claim to have knowledge about topics that actually do not exist. This tendency is especially pronounced in areas in which people claim to be highly knowledgeable (Atir, Rosenzweig, & Dunning, 2015). For example, participants who rated themselves to be knowledgeable about personal finance were more likely to claim knowledge about fictitious financial constructs, such as "fixed-rate deduction." Another example is the *illusion of explanatory depth*, in which people believe that they understand how common objects, such as toilets and toasters, work, when in fact they do not (Fernbach, Rogers, Fox, & Sloman, 2013; Rozenblit & Keil, 2002). It appears that our familiarity and experience with these objects creates a misplaced sense of understanding. Dunning (2011) describes numerous instances of situations, ranging from the trivial to the calamitous, in which

people demonstrate that they are unaware of how little they know or how poorly they understand a topic or situation, leading them to sometimes dangerous actions (e.g., rewiring the electrical system in one's house). In such cases, people have a metacognitive experience feeling that leads them to believe that they know more than they actually do.

A third example is the *Dunning–Kruger effect*, in which people demonstrate more confidence in their responses than their performance warrants (Kruger & Dunning, 1999; see also Fischhoff, Slovic, & Lichtenstein, 1977). Of particular relevance to the rationality debate is the demonstration that this effect extends to tasks designed to measure reasoning biases (Pennycook, Ross, Koehler, & Fugelsang, 2017). In one study, participants completed the *cognitive reflection test* (Frederick, 2005), which consists of three items such as the following:

> A ball and a bat together cost $1.10. The bat costs $1.00 more than the ball. How much does the ball cost?

About two-thirds of respondents mistakenly answer "10¢," when a bit of basic algebra would indicate that the correct answer is "5¢." The answer "10¢" comes easily to mind, possibly because people miss or misinterpret the "more than" (Mata, Ferreira, Voss, & Kollei, 2017). As such, the results of this test are commonly interpreted as a measure of people's tendency to respond intuitively rather than reflectively (Toplak, West, & Stanovich, 2011). Pennycook et al. (2017) observed a typical Dunning–Kruger effect: those who performed well tended to underestimate their performance, whereas those who performed most poorly believed they did three times better than they actually did (for related findings, see Mata, Ferreira, & Sherman, 2013). The poor performers are therefore "doubly cursed" (Kruger & Dunning, 1999): not only do they perform poorly, but they are unaware of it and therefore are not in a position to ameliorate their performance.

There are two other relevant points to make about confidence and overconfidence. First, evidence has emerged showing that confidence has a trait-like property, in that people who are confident (or overconfident) about their performance in one domain also tend to be so in other domains (Jackson & Kleitman, 2014; Stankov, Kleitman, & Jackson, 2014). Second, as argued above, confidence is the arbiter of action. In this case, confidence has been found to predict decision-making style (Jackson & Kleitman, 2014; Jackson, Kleitman, Stankov, & Howie, 2016). To test this hypothesis, reasoners were asked whether they wanted to take action on each decision, for example, by submitting an answer (i.e., to count as part of the total score) or administering a treatment for a fictitious disease. It was found that confident reasoners take actions that are congruent with the decision they made, regardless of whether it was accurate. Consequently, those who are overconfident make errors of commission (i.e., they act when they should not). Conversely, those who are underconfident make errors of omission (i.e., fail to act when they are correct).

Up to this point, we have discussed ways in which metacognition is a source of epistemic irrationality, by producing monitoring judgments, based on unreliable cues, that lead to overconfidence. A more optimistic view comes from work on consistency and on metacognitive coherence. For the latter, read: regularity and predictability; it means the sense of fit between environmental cues and our knowledge and expectations (Topolinski, 2015, 2018).[3]

We start with work showing that metacognitive judgments are sensitive to semantic coherence, that is, they are sensitive to the rules that govern associations between entries in our mental lexicon[4] (Sweklej, Balas, Pochwatko, & Godlewska, 2014; for a related view, see Betsch & Glöckner, 2010). Much of the work in support of this claim is based on a variation of Mednick and Mednick's (1967) Remote Associates Test. The experimental paradigm involves presenting participants with word triads. The researchers defined triads to be coherent when the three words were associated with a fourth word (e.g., "struck," "beam," and "light" are associated with "moon"). For incoherent triads, there is no common remote associate (e.g., "car," "tiger," and "cream"). Semantic coherence, in this context, is a type of fluency cue—it triggers a feeling of processing ease (or, if lacking, difficulty). Bolte and Goschke (2005) observed that participants are able to identify semantic coherence of triads at above-chance rates, even when they were required to make such judgments after a brief presentation (namely, in 1.5 seconds; for a review, see Topolinski, 2015; but see Ackerman & Beller, 2017).

Although it is not usual to think of them in such a way, the judgments of semantic coherence used in this line of research are akin to metacognitive monitoring judgments done regarding one's own cognitive performance (Thompson, 2014). We know that people can make judgments regarding their knowledge under similar time constraints (Reder & Ritter, 1992). Moreover, both types of judgments make an inference about a mental state, based on the heuristic cues that generate feeling of fluency (Topolinski, 2015). In Ackerman and Thompson's (2015, 2017) framework, both are similar to judgments of solvability, which, as described above, are prospective judgments about whether the participant would be able

to solve the problem at hand (see examples in Ackerman & Beller, 2017; Lauterman & Ackerman, 2019).

While consistency and coherence are different aspects of metacognitive judgment, they both have to do with regularity and predictability versus conflict. Evidence for the role of consistency in metacognitive judgments can be found in the so-called *conflict detection effect* (De Neys, 2014). Many classic reasoning problems are designed to be tricky, in that they pit two conflicting answers against each other, as in the bat and ball problem above. People's confidence in their answers to such problems is lower than it is to control versions in which there is no trick (De Neys, Cromheeke, & Osman, 2011; Thompson et al., 2011; Thompson & Johnson, 2014). That is, when such problems generate a feeling of disfluent processing, by permitting two conflicting answers, it is reflected in people's confidence judgments.

Indeed, Koriat (2012) argued that consistency is the major determinant of confidence and also underlies fluency. His *self-consistency model* (SCM) defines self-consistency as a mnemonic cue that captures agreement among a variety of considerations, including task performance and knowledge. The SCM applies in cases where there are two alternatives from which to choose. To decide, people are assumed to gather/retrieve information about those two alternatives. Confidence in the decision is determined by the consistency of the information that favors the preferred alternative: choices that have many pieces of information favoring them, and few drawbacks, engender confidence and will be made relatively more fluently than those that are less consistent.

3. Normative, Practical, and Bounded Rationality in Light of Metacognitive Research

So far, the literature we reviewed resonates well with the *heuristics and biases* tradition in the psychology of decision making (e.g., Gilovich, Griffin, & Kahneman, 2002; and see chapter 2.4 by Fiedler, Prager, & McCaughey, this handbook). The core idea of heuristics and biases is *meliorist* (a term we borrow from Stanovich, 1999): people draw on heuristic mental shortcuts, which in everyday life often work well enough, but in many contexts can be spectacularly off the mark, leading to biased reasoning and decision making. This approach highlights the notorious *normative–descriptive gap*: it holds that there is a set of norms that ought to be met (such as those of the probability calculus or classical logic), but actual behavior falls short of them.

However, there is more to rationality than meeting normative rules (or "rationality$_2$," as Evans & Over, 1996, labeled it). Rationality can also be gauged by practical or instrumental standards ("rationality$_1$," in the Evans and Over nomenclature): does the thought or decision or act lead to achieving one's goals? Behavior that is normatively irrational can often be argued to be pragmatically rational and vice versa. The term often mentioned in this context is "bounded rationality" (Simon, 1982; and see chapter 8.5 by Hertwig & Kozyreva, this handbook), an idea that turns out to be surprisingly metacognitive. Simon highlights the inherent cognitive and biological limitations on human thinking: we do not have unbounded working memory or unlimited attentional resources; we do not live forever. Some normative solutions to rational puzzles are computationally intractable in that they would require more than the lifetime of the universe to be computed, never mind a human life span. Thus, with human rationality necessarily bounded by these limitations, the rational approach is, rather than to try and find the slippery, often intractable, optimal solutions, to scale down to solutions that are just good enough: to *satisfice*, in Simon's terminology.

Seen through the lens of metacognitive research, satisficing is a type of stopping rule (see Ackerman & Thompson, 2015, 2017): it tells us when we can stop searching for a better response, because the one we have is good enough. Nonetheless, the concept of satisficing on its own is rather underspecified. All it tells us is that people abort the search prior to achieving a normatively valid solution, but it does not tell us the details of the underlying psychological mechanisms, for example, that conceptual fluency (ease of processing) might trigger aborting the search regardless of how close to optimal the solution is. The unique insight from metacognitive research is that satisficing is not some sort of downgraded normative processing, where stopping depends on some quantitatively partial fulfilment of normative parameters. Rather, stopping processing depends on cues that are *qualitatively different* from normative parameters of the solution and only happen to (modestly) correlate with them (see Ackerman, 2014).

4. Meta-Reasoning and the Rationality of Persistence

4.1 Are Humans Cognitive Misers?

One surprising insight that metacognitive research on reasoning and decision making can afford us on human rationality regards the so-called *cognitive miser hypothesis*. Originally proposed in the context of social cognition (Fiske & Taylor, 1984, 1991), the cognitive miser hypothesis suggests that people are mental Harpagons, stingily doling out cognitive resources like the miser in

Molière's play reluctantly giving up his gold. For example, people are capable of overcoming social stereotyping to treat people as individuals rather than cast types, but this demands mental effort. Although people possess the requisite mental resources, they are often loath to invest them, thus succumbing to stereotyping due to mental laziness. The cognitive miser hypothesis was later adopted into the rationality debate by the meliorist tradition (Evans & Stanovich, 2013; Stanovich, 2009), where it transmuted into the more specific idea that biases (and therefore irrationality) result from defaulting into less effortful processing.

How sound is the cognitive miser hypothesis? There are several presuppositions here: (1) people tend to stop searching for solutions prematurely, even though (2) they have the necessary cognitive resources to complete the search, and (3) this leads to bias and normative irrationality. It follows that (4) investing more cognitive effort should correlate positively with being normatively rational. The problem is that each one of these theses is suspect. Theses (2) and (3) have both been challenged before (for thesis (2), see, e.g., Elqayam, 2012; for thesis (3), see Gigerenzer, Todd, & ABC Research Group, 1999). Where meta-reasoning research can provide novel insight is mainly with (1) and (4). Contrary to widespread consensus, people often *fail* to satisfice, and any correlation between cognitive investment and normative responding is weak at best (for a review, see Ackerman, 2014).

4.2 Thinking in Vain

Bounded rationality and satisficing have long been a source of contention within the great rationality debate. Within the heuristics and biases (aka meliorist) tradition, the term "bounded rationality" is used almost as a synonym for "error." In contrast, the Panglossian tradition (again borrowing a term from Stanovich, 1999), particularly in the literature dealing with the fast-and-frugal heuristics approach (Gigerenzer et al., 1999), sees bounded rationality as the epitome of human rationality (see also chapter 8.5 by Hertwig & Kozyreva, this handbook). In this chapter, we take no side in that particular debate, except to note that both sides share the presupposition that people do indeed satisfice—which is exactly what metacognitive research calls into question. There are numerous cases, we learn from metacognitive research, in which the opposite is true: people systematically *waste* cognitive effort.

Nelson and Leonesio (1988) called this *labor in vain*. Metacognitive research shows that when people can self-pace their own learning or problem solving, they invest the longest time in the most difficult items, those for which they have little chance of success (Ackerman, 2014; Koriat, Ma'ayan, & Nussinson, 2006; Undorf & Ackerman, 2017). They do so even when they themselves judge their chances of success with these items to be slim and even when they are offered incentives for accuracy (in fact, incentives for accuracy exacerbate, rather than mitigate, labor in vain). These participants over- rather than underinvest cognitive effort, making them cognitive *wasters* rather than cognitive *misers*. The labor-in-vain hypothesis is diametrically opposed to the cognitive miser hypothesis. Importantly, the weight of the metacognitive evidence favors labor in vain.

A major source of evidence comes from Ackerman's (2014) *diminishing criterion model* (DCM). It shows that although reasoners do adjust their levels of aspiration—and, as a consequence, their rate of satisficing—to changing circumstances, the updating lags far behind diminishing resources. In Ackerman's (2014) experiment 3, for example, participants solved the easiest items in about 10 seconds with over 90% confidence, whereas the most difficult items took about 90 seconds, and even then, participants' confidence was not much higher than 20%. Thus, it took a long time for the penny to drop and for participants to stop trying for a better (i.e., more confidently held) solution. An important component in the DCM is the time limit—the maximum amount of time the respondent is willing to invest in each task item. Recent studies have shown across a large variety of tasks, including memorizing, various problem-solving tasks, and real-life web searches, that people limit the amount of time that they spend considering their response, with the aim of not wasting time on items they expect to make little progress on. However, it is nonetheless the case that the items on which the most time is spent have the lowest rates of success (Ackerman, Yom-Tov, & Torgovitsky, 2020; Undorf & Ackerman, 2017). Thus, it is not always the case that reasoners resort to satisficing, and even when they do, they sometimes fail to satisfice sufficiently, even by their own lights, when they judge their chances of success to be slim.

The labor-in-vain effect has also been observed in people's information-gathering strategies. This research takes place in the fast-and-frugal heuristics research tradition, where it is posited that people solve complex problems using simple heuristics (Gigerenzer et al., 1999). Nonetheless, it has been observed that to make a decision, people often gather more information than specified by the heuristic. A common task asks people to decide between a pair of alternatives, such as which stock to invest in. They are allowed to gather information about each of the options prior to reaching a decision. People

often continue to gather information even when they already have enough information to discriminate between options (see Bröder & Newell, 2008; for reviews, see Hilbig, 2010). Indeed, people will pay for that information, even if it is not helpful (Newell, Weston, & Shanks, 2003). This suggests that having additional information, regardless of quality, may help people reach a confidence threshold where they are comfortable with their decision.

4.3 Cognitive Investment and Normative Responding

We now turn to the idea that there should be a positive relation between investment of cognitive effort and normatively correct responding. The main evidence against this idea comes from Thompson's two-response paradigm (Thompson et al., 2011). The core method in this paradigm consists of presenting participants with (typically) classic reasoning, decision-making, or moral judgment problems and asking them to provide a first quick, intuitive response, followed by a second, more considered response. Each of the two responses is also accompanied by rating on a metacognitive confidence scale. For the first response, the judgment is the feeling of rightness (FOR) mentioned above. For the second response, the judgment reflects a final judgment of confidence. The typical finding is that people seldom change their response from the first to the second response, but they are more likely to do so when metacognitive cues trigger low FOR—typically as a result of processing disfluency. As noted above, these cues only marginally coincide with the normatively correct response.

The cognitive miser hypothesis would suggest that in the rare cases that people change their mind between the first and the second response, it should be more often in the direction of a normatively correct response than the opposite. Actual findings suggest that the connection is much more complex. For example, people change from a normative to a nonnormative solution almost as often as the other way around. While there often is an increase in normatively correct responses over time, it is typically small (e.g., Bago & De Neys, 2019; Shynkaruk & Thompson, 2006; Thompson et al., 2011; Thompson & Johnson, 2014). Thus, investing more time and thought does not necessarily lead to more normative solutions.

Similarly, Stupple, Pitchford, Ball, Hunt, and Steel (2017) found only a small positive correlation between latencies (a potential proxy for cognitive effort) and correct solutions for the cognitive reflection test (Frederick, 2005), which aims to measure the ability to provide normatively correct responses against the lure of intuitively compelling but incorrect responses. Even more interesting are item-level results. For the bat and ball problem (which requires mental arithmetic to solve correctly), the cognitive miser hypothesis was supported: latencies positively predicted the percentage of correct responses, and correct responses took longest. However, on the lily pad problem (which requires little mental arithmetic and is more in the nature of an insight item), latencies actually *negatively* predicted normatively correct responses. For the latter, intuitively compelling erroneous responses were fastest, but nonintuitive erroneous responses were slowest—slower than correct responses, in fact. This suggests that reasoners were investing time and effort, just to end up with the wrong answers.

Taken together, the findings from Thompson et al. (2011) and from Stupple et al. (2017) seem to hint at an inverted-U-shaped relation between normative correctness and cognitive resources. Beyond a certain point, additional cognitive resources have little role to play and might in fact turn out to be detrimental to normative responding. If this is correct, we can outline the boundaries of the cognitive miser hypothesis, which seems to have limited scope; beyond that, evidence supports the labor-in-vain hypothesis.

4.4 Rationality and Further Deliberation

Last, we turn to the contribution of meta-reasoning research to a little-explored but significant issue in rationality: information gathering. Classic Bayesian decision theory provides a normative model for selecting between alternative options but is silent on how an option set is assembled in the first place. When do we stop looking for alternatives? Suppose you are looking for a second medical opinion. Does it make sense to look for a third opinion? A fourth? Bounded rationality tells us that eventually people will abort the search and that they are likely to do so prematurely, but this principle is underspecified. Douven (2002) outlined a philosophical model of the rationality of further deliberation, suggesting that in any given search, the rationality of searching further is determined by four parameters:

1. How satisfactory does the agent find the existing options? Ceteris paribus, the better the existing options, the lower the value of further search.
2. What is the agent's estimate, or prediction, of her chance of success in finding a better alternative? Ceteris paribus, the stronger the prediction of success, the higher the value of further search.
3. What is the cost of further search? Ceteris paribus, the higher the cost, the lower the rationality of further search.
4. What is the time limit for the search?

These options can all be operationalized in metacognitive terms (Ackerman, Douven, Elqayam, & Teodorescu, 2020). The first thing to note is that parameter (1) is closely related to Thompson's feeling of rightness, the metacognitive confidence measure. The higher the FOR, the more satisfactory the agent will find the existing options and the less inclined they will be to search further. Evidence for this comes from the two-response paradigm and from reasoners' reluctance to change responses accompanied by high FOR. Parameter (2), prediction of success, might be akin to judgment of solvability (Ackerman & Thompson, 2017).

An important insight that comes from meta-reasoning research is that parameters (3) and (4)—that is, the cost of further search and time limit—are not independent: in Ackerman's (2014) diminishing criterion model, time itself is a cost: as time goes by, participants are increasingly willing to settle for less ambitious goals, in which they have lower confidence.

5. Conclusions

In sum, we have argued that understanding metacognitive processes can advance our understanding of epistemic, pragmatic, and bounded rationality. Because cues to confidence may be unrelated to the accuracy or adequacy of decisions, we are often overconfident in our reasoning performance and lay claim to knowledge about fictitious concepts. We recast Simon's (1982) notion of satisficing as a metacognitive stopping rule, demonstrating that in many circumstances, people fail to satisfice, even when they have enough information to make an informed choice. Thus, rather than being cognitive misers, people tend to labor in vain, spending precious cognitive resources on difficult or unsolvable problems.

Notes

The first two authors contributed equally to the preparation of the chapter.

1. Nelson and Narens (1990) borrowed the "meta" from Hilbert's "metamathematics" and Carnap's "metalanguage." Nelson (1996) anchors the term in Tarski's philosophy of truth, where "object language" denotes referents that are not linguistic (e.g., chair, happiness), while "meta-language" (and "meta-meta-language," and so on) refers to linguistic terms such as "truth." Thus, "Snow is white" is in the object language, whereas "It is true that snow is white" is in the meta-language. The "meta" in "metacognition" similarly refers to a referential level of processing: the content of thought is the object level and the monitoring of that content is the meta-level.

2. In philosophy, the role of truth-conduciveness in epistemic rationality is moot, particularly the relation between epistemic and practical rationality. As Stich (1999) argued, the conception of truth in philosophy is too fragmented to allow a single normative standard. In psychology of reasoning, Evans (2014) advocated a view of epistemic rationality as subservient to pragmatic rationality. For a recent defense of the connection between truth-conduciveness and pragmatic rationality, see Schurz (2014) and Schurz and Hertwig (2019), although we note that Schurz concedes that this view is somewhat undermined by placebo effects (beneficial false beliefs). Here we stick to the classic notion of epistemic rationality based on truth-conduciveness, as this is the view that dominates most of the metacognitive literature, but the reader is invited to keep this caveat in mind.

3. The metacognitive conception of coherence is therefore distinct from the Bayesian one in than it does not rely on conformity to the probability calculus.

4. The term "semantic coherence" is borrowed from linguistics, where it means the extent to which words adhere to a word formation rule, so that they are predictable given the rule (Aronoff, 1976, p. 38).

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General*, *143*(3), 1349–1368.

Ackerman, R. (2019). Heuristic cues for meta-reasoning judgments: Review and methodology. *Psychological Topics*, *28*(1), 1–20.

Ackerman, R., & Beller, Y. (2017). Shared and distinct cue utilization for metacognitive judgements during reasoning and memorisation. *Thinking & Reasoning*, *23*(4), 376–408.

Ackerman, R., Douven, I., Elqayam, S., & Teodorescu, K. (2020). Satisficing, meta-reasoning, and the rationality of further deliberation. In S. Elqayam, I. Douven, J. St. B. T. Evans, & N. Cruz (Eds.), *Logic and uncertainty in the human mind: A tribute to David Over* (pp. 10–26). London, England: Routledge.

Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 164–178). London, England: Psychology Press.

Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Science*, *21*(8), 607–617.

Ackerman, R., Yom-Tov, E., & Torgovitsky, I. (2020). Using confidence and consensuality to predict time invested in problem solving and in real-life web searching. *Cognition*. Advance online publication. doi.org/10.1016/j.cognition.2020.104248

Ackerman, R., & Zalmanov, H. (2012). The persistence of the fluency–confidence association in problem-solving. *Psychonomic Bulletin and Review*, *19*, 1187–1192.

Aronoff, M. (1976). *Word formation in generative grammar*. Cambridge, MA: MIT Press.

Atir, S., Rosenzweig, E., & Dunning (2015). When knowledge knows no bound: Self-perceived expertise predicts claims of impossible knowledge. *Psychological Science*, *26*(8), 1295–1303.

Bago, B., & De Neys, W. (2019). The smart system 1: Evidence for the intuitive nature of correct responding on the bat-and-ball problem. *Thinking & Reasoning*, *25*(3), 257–299.

Betsch, T., & Glöckner, A. (2010). Intuition in judgment and decision making: Extensive thinking without effort. *Psychological Inquiry*, *21*, 1–16.

Bolte, A., & Goschke, T. (2005). On the speed of intuition: Intuitive judgments of semantic coherence under different response deadlines. *Memory & Cognition*, *33*(7), 1248–1255.

Bröder, A., & Newell, B. R. (2008). Challenging some common beliefs: Empirical work within the adaptive toolbox metaphor. *Judgment and Decision Making*, *3*, 205–214.

De Neys, W. (2014). Conflict detection, dual processes, and logical intuitions: Some clarifications. *Thinking & Reasoning*, *20*, 169–187.

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, *6*, e15954.

Douven, I. (2002). Decision theory and the rationality of further deliberation. *Economics & Philosophy*, *18*(2), 303–328.

Dunning, D. (2011). The Dunning–Kruger effect: On being ignorant of one's own ignorance. In J. Olson & M. P. Zanna (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 247–296). New York, NY: Elsevier.

Elqayam, S. (2012). Grounded rationality: Descriptivism in epistemic context. *Synthese*, *189*(1), 39–49.

Evans, J. St. B. T. (2014). Two minds rationality. *Thinking & Reasoning*, *20*(2), 129–146.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, England: Psychology Press.

Evans, J. St. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223–241.

Fernbach, P. M., Rogers, T., Fox, C. R., & Sloman, S. A. (2013). Political extremism is supported by an illusion of understanding. *Psychological Science*, *24*, 939–946.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*, 552–564.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. New York, NY: McGraw-Hill.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition: From brains to culture* (2nd ed.). Los Angeles, CA: Sage.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

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

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgement*. Cambridge, England: Cambridge University Press.

Hilbig, B. E. (2010). Reconsidering "evidence" for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, *17*, 923–930.

Jackson, S. A., & Kleitman, S. (2014). Individual differences in decision-making and confidence: Capturing decision tendencies in a fictitious medical test. *Metacognition and Learning*, *9*, 25–49.

Jackson, S. A., Kleitman, S., Stankov, L., & Howie, P. (2016). Individual differences in decision making depend on cognitive abilities, monitoring and control. *Journal of Behavioral Decision Making*, *30*, 209–223.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370.

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–325). Cambridge, England: Cambridge University Press.

Koriat, A. (2012). The self-consistency model of subjective confidence. *Psychological Review*, *119*(1), 80–113.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General*, *135*(1), 36–69.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134.

Lauterman, T., & Ackerman, R. (2019). Initial judgment of solvability in non-verbal problems—a predictor of solving processes. *Metacognition and Learning, 14*(3), 365–383.

Mata, A., Ferreira, M. B., & Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process

perspective on overconfidence. *Journal of Personality and Social Psychology*, *105*(3), 353–373.

Mata, A., Ferreira, M. B., Voss, A., & Kollei, T. (2017). Seeing the conflict: An attentional account of reasoning errors. *Psychonomic Bulletin & Review*, *24*, 1980–1986.

Mednick, S. A., & Mednick, M. T. S. (1967). *Examiner's manual, Remote Associates Test: College and adult forms 1 and 2*. Boston, MA: Houghton Mifflin.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102–116.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press.

Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best." *Organizational Behavior and Human Decision Processes*, *91*, 82–96.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2017). Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence. *Psychonomic Bulletin & Review*, *24*, 1474–1484.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 435–451.

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Schurz, G. (2014). Cognitive success: Instrumental justifications of normative systems of reasoning. *Frontiers in Psychology*, *5*. doi:10.3389/fpsyg.2014.00625

Schurz, G., & Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Topics in Cognitive Science*, *11*(1), 7–36.

Shynkaruk, J. M., & Thompson, V. A. (2006). Confidence and accuracy in deductive reasoning. *Memory & Cognition*, *34*, 619–632.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Stankov, L., Kleitman, S., & Jackson, S. A. (2014). Measures of the trait of confidence. In G. J. Boyle, D. H. Saklofske, & G. Matthews (Eds.), *Measures of personality and social psychological constructs* (pp. 158–189). Amsterdam, Netherlands: Academic Press.

Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.

Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In J. St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford, England: Oxford University Press.

Stich, S. P. (1990). *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation*. Cambridge, MA: MIT Press.

Stupple, E. J. N., Pitchford, M., Ball, L. J., Hunt, T. E., & Steel, R. (2017). Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the Cognitive Reflection Test. *PLoS ONE*, *12*(11), e0186404.

Sweklej, J., Balas, R., Pochwatko, G., & Godlewska, M. (2014). Intuitive (in)coherence judgments are guided by processing fluency, mood, and affect. *Psychological Research*, *78*, 141–149.

Thompson, V. A. (2014). What intuitions are . . . and are not. *Psychology of Learning and Motivation*, *60*, 35–75.

Thompson, V. A. (2016). Certainty and action. In N. Galbraith, E. Lucas, & D. Over (Eds.), *The thinking mind: A Festschrift for Ken Manktelow* (pp. 80–96). Oxford, England: Routledge.

Thompson, V. A., Evans, J. St. B. T., & Campbell, J. I. C. (2013). Matching bias on the selection task: It's fast and it feels good. *Thinking & Reasoning*, *19*, 431–452.

Thompson, V. A., & Johnson, S. C. (2014). Conflict, metacognition, and analytic thinking. *Thinking & Reasoning*, *20*(2), 215–244.

Thompson, V. A., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind & Society*, *11*, 93–105.

Thompson, V. A., Prowse Turner, J., & Pennycook, G. (2011). Intuition, metacognition, and reason. *Cognitive Psychology*, *63*, 107–140.

Toplak, M., West, R. F., & Stanovich, K. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, *39*, 1275–1289.

Topolinski, S. (2015). Intuition: Introducing affect into cognition. In A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 146–163). London, England: Psychology Press.

Topolinski, S. (2018). The sense of coherence: How intuition guides reasoning and thinking. In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking and reasoning* (pp. 559–574). London, England: Routledge.

Undorf, M., & Ackerman, R. (2017). The puzzle of study time allocation for the most challenging items. *Psychonomic Bulletin & Review*, *24*(6), 2003–2011.

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