19 Metacognitive illusions

Monika Undorf, Sofia Navarro-Báez, and Malte F. Zimdahl

People often reflect on their own learning, memory, and thinking. Imagine, for instance, students who are preparing for an exam. They engage in study activities such as reviewing lecture slides, rereading book chapters, writing summaries, or using flashcards. At the same time, they think about whether they understand the texts they are reading, consider how well they can remember definitions, concepts, and theories, and evaluate whether they have sufficiently learned the material to succeed in the exam. Thus, some of the students’ cognitions are cognitive processes about cognitive processes, often termed metacognition (Flavell, 1971; Nelson & Narens, 1990).

Metacognition entails two components (Nelson & Narens, 1990). Metacognitive monitoring refers to thoughts, knowledge, and judgments about cognitive processes as well as to assessments of one’s own cognitions. Metacognitive control refers to the use of this information for regulating cognition and behavior. In our example, the students engage in metacognitive monitoring when thinking about their understanding and learning, when reflecting on their learning progress, and when assessing their overall level of knowledge. They engage in metacognitive control when utilizing the output of their monitoring processes to self-regulate their learning. A student who feels that he does not progress well might try a different study strategy or take a break, whereas a student who thinks that she has sufficiently mastered the material might stop studying altogether.

Decades of research have demonstrated that accurate metacognition is critical for good performance on various cognitive tasks. For instance, a recent meta-analysis showed that accurate metacognition positively predicts academic performance in adults, adolescents, and children even when controlling for intelligence (Ohtani & Hisasaka, 2018). Thus, much can be gained from accurate metacognition. At the same time, there is a real danger that metacognitive illusions undermine cognitive performance.

Metacognitive illusions are defined as reliable and systematic dissociations between people’s metacognitions and cognitions. Thus, unlike optical or cognitive illusions, illusory metacognitions do not deviate from some external “reality” (see Introduction) but from the cognitions they are supposed to assess. Apart from this difference, however, cognitive and metacognitive illusions share their defining features. In particular, illusory metacognitions occur involuntarily, clash with people’s conviction that they know their own minds, are difficult to avoid, and have attracted a great deal of interest from researchers and practitioners (e.g., Dunlosky & Metcalfe, 2009; Koriat, 2007; Undorf, 2020).

In this chapter, we present an overview of metacognitive illusions and describe how to explore metacognitive illusions in the classroom. We then discuss theoretical accounts for metacognitive illusions, point out real-life consequences of metacognitive illusions, and describe research on the mending of metacognitive illusions.
Text box 19.1 Metacognitive illusions in real life

1) An employee who is faced with a logical problem works on it alone rather than together with her colleagues, because she believes that individual workers achieve as least as good reasoning performance as work teams. This metacognitive belief is held by people from various countries including managers and psychologists who study reasoning. However, it is in stark contrast to evidence that group discussions considerably improve reasoning performance (Mercier et al., 2015).

2) A responsible citizen aims to learn more about the positions of parliamentary candidates regarding various political topics. Because she considers herself well informed about the most relevant and most discussed issues, she focuses on topics that have not been much in the news lately. Research suggests that she falls prey to a metacognitive illusion: People overestimate their understanding of political topics they feel everyone should know about (Gaviria & Corredor, 2021).

The assessment of metacognition

Metacognitive judgments

Researchers assess metacognitive monitoring by asking people to judge their own cognitions. Numerous metacognitive judgments have been obtained (e.g., Dunlosky & Metcalfe, 2009). Among the most popular metacognitive judgments are the following: Judgments of learning (JOLs) are made during learning and predict one’s chances of remembering recently studied information on a later test. Feeling of knowing (FOK) judgments predict the likelihood that currently unrecallable information is nevertheless available in memory and will be remembered in the future. Metacomprehension judgments are people’s judgments about how well they have comprehended and learned text materials. People’s judgments of solvability assess whether they can solve specific problems. Confidence judgments assess people’s confidence in the correctness of their responses to cognitive tasks.

Accuracy of metacognitive judgments

Metacognitive judgments are accurate if they correspond well with cognitive performance. Because high correspondence may take different forms, three aspects of metacognitive accuracy need to be distinguished. The first aspect is calibration (or absolute accuracy; e.g., Koriat, 2007). Metacognitive judgments are well calibrated if their level is similar to the level of actual performance. If a person’s mean confidence judgment in her responses is 60% and she gives correct answers to six out of ten items, then absolute accuracy is excellent. Alternatively, metacognitive judgments can be overconfident or underconfident. Overconfident metacognitive judgments significantly exceed cognitive performance (see Chapter 18), whereas underconfident metacognitive judgments significantly underestimate cognitive performance. A widely used measure of calibration is the signed difference between a person’s mean metacognitive judgments and the person’s mean performance, referred to as bias in studies on metacognition (e.g., Dunlosky & Metcalfe, 2009) and as over-/underconfidence in Chapter 18. A positive bias reflects overconfidence,
a negative bias reflects underconfidence, and a bias near zero reflects good calibration. Calibration is relevant for metacognitive control processes at the task level. For instance, well-calibrated JOLs allow learners to devote adequate time and effort to a learning task as a whole. In contrast, overconfident learners prematurely terminate studying and receive disappointing test grades, whereas underconfident learners spend too much time and effort (Koriat, 2007).

The second aspect of metacognitive accuracy is resolution (or relative accuracy; e.g., Koriat, 2007). Resolution refers to the extent to which metacognitive judgments distinguish between correct and incorrect responses. If a person displays high confidence in all correct responses and low confidence in all incorrect responses, then relative accuracy is excellent. A widely used measure of resolution is the within-person Goodman-Kruskal gamma correlation between judgments and performance (Nelson, 1984; for criticism and alternatives, see, e.g., Bröder & Undorf, 2019). Gamma can range from $-1$ to $1$, with higher values indicating higher resolution than lower values. Resolution is relevant to metacognitive control processes at the item level. High resolution of JOLs helps learners to selectively allocate study time and effort to material that is not yet well learned. In contrast, low resolution of JOLs may result in learners allocating study time and effort to material they have already mastered (Koriat, 2007). It is important to note that calibration and resolution are largely independent. Metacognitive judgments can be grossly overconfident but nevertheless exhibit high resolution. Also, metacognitive judgments can be well calibrated but still have poor resolution.

Finally, metacognitive judgments can be accurate in that they track differences in actual performance across conditions. This means that metacognitive judgments and actual performance show the same pattern across conditions. If a person is more confident in her responses on a music quiz than an astronomy quiz and indeed achieves better performance on the music quiz, then metacognitive judgments accurately track performance differences across quizzes. Notably, this does not necessarily imply that the person’s confidence judgments are well-calibrated (e.g., confidence might exceed performance on both quizzes) or reveal high resolution (e.g., confidence might be similar for correct and incorrect responses in one or both quizzes). Accurate tracking of performance differences across conditions is relevant for metacognitive control processes at an intermediate level. Recognizing that a certain subject, topic, or type of problem is particularly difficult allows learners to allocate more study time and effort to this subject, topic, or problem type than to others.

Because the remainder of this chapter focuses on metacognitive illusions, it must be emphasized that metacognitive judgments are often quite accurate in terms of calibration and resolution and usually track performance differences across conditions (e.g., Koriat, 2007). Thus, as is the case with cognitive illusions (see Chapter 1), metacognitive illusions are the exception rather than the rule. Also, some metacognitive illusions were found in rather artificial experimental situations (see, e.g., Undorf, 2020). Finally, metacognitive illusions notwithstanding, metacognitive processes are of considerable adaptive and functional value (e.g., Benjamin, 2007).

Overview of metacognitive illusions

This section presents an overview of illusions that afflict judgments about one’s own learning, knowledge, and thinking (see Figure 19.1).
Metacognitive illusions of learning

A well-replicated metacognitive illusion that impairs the calibration of predictions about learning is the *underconfidence-with-practice effect* (Koriat, 1997; see also Koriat et al., 2002). Koriat (1997) discovered that the calibration of JOLs deteriorates when people study and recall the same information across multiple study-test trials. Each study-test trial comprised a study phase in which people studied items such as word pairs (e.g., *glass – island*) and made JOLs, that is, predicted their chances of recalling each second word when seeing the first word at test. JOLs showed good calibration on the first study-test trial but shifted towards marked underconfidence from the second trial onwards. The reason for this was that people underestimated just how much memory performance benefits from study-test practice. Importantly, however, detrimental effects of study-test practice on JOL accuracy are limited to calibration. Repeated study-test trials typically improve JOL resolution, that is, increase the extent to which the JOLs differentiate between items that are remembered and items that are not remembered.

The *stability bias* (Koriat et al., 2004; Kornell & Bjork, 2009) is a metacognitive illusion in which people fail to anticipate future learning or future forgetting. Kornell and Bjork (2009) examined the impact of future study opportunities on JOLs. They asked participants to predict their chances of remembering recently studied items in tests that were announced to follow immediately upon the study phase or to take place after one, two, or three additional study-test trials. As expected, studying resulted in learning: Memory performance increased by 36% from Test 1 to Test 4. JOLs, in contrast, showed an unreliable increase of only 8%. The stability bias thus produced increasingly underconfident JOLs that failed to track performance improvements resulting from future study-test
trials. Koriat et al. (2004) showed that JOLs were insensitive to future forgetting. In their study, three groups of participants made JOLs for a memory test that was announced to take place immediately after studying, one day later, or one week later. Unsurprisingly, actual memory performance declined with longer retention intervals: Percentage correct dropped from 53% (immediate test) to 18% (test after one week). JOLs, however, were completely indifferent to the expected retention interval (immediate test: 52%; test after one week: 54%) and, consequently, severely overconfident with longer retention intervals. Here, the instability bias produced overconfident JOLs that failed to track performance impairments due to future forgetting.

A metacognitive illusion with dramatic effects on resolution is the semantic-fluency illusion (Benjamin et al., 1998). In Benjamin et al.'s (1998) experiment, participants answered general-knowledge questions and predicted the likelihood of recalling their answers on a free recall test in a second phase of the experiment. While actual free recall performance increased with the time it took people to answer the general-knowledge questions in Phase 1, JOLs decreased with response time in Phase 1. This dissociation resulted in negative gamma correlations between JOLs and free recall performance, indicating that people made lower JOLs for items they remembered than for items they did not remember. A related metacognitive illusion is the search-fluency illusion (Stone & Storm, 2021): People who used the internet to search for answers to general-knowledge questions made higher JOLs for the free recall of answers they found quickly rather than slowly, even though recall performance increased with search time.

The font-size illusion (Rhodes & Castel, 2008; for a review, see Luna et al., 2018) impairs the tracking of performance differences in JOLs. This illusion occurs when people make JOLs for study words that are printed in different font sizes (e.g., in a larger 48-point font and in a smaller 18-point font). People provide higher JOLs for large-font words than for small-font words, even though large-font words are remembered only slightly better than small-font words, if at all. Calibration and resolution typically remain unaffected by the font-size illusion.

Metacognitive illusions also afflict people’s JOLs for complex study materials. An example is the instructor-fluency illusion (Carpenter et al., 2013). In Carpenter et al.’s (2013) study, students viewed one of two videotaped lectures. In the fluent video, the lecturer spoke freely and fluently and maintained eye contact with the camera. In the disfluent video, the same lecturer read the same content haltingly from notes, repeatedly flipped through her notes, and switched her gaze between camera and notes. Participants who viewed the fluent video predicted that they would remember a greater amount of information than those who viewed the disfluent video (48% v. 25%), even though actual memory performance was similar across the two groups (26% v. 22%). Viewing the fluent video thus led to severely overconfident JOLs.

The calibration of metacomprehension judgments predicting people’s memory for texts is afflicted by screen inferiority (Ackerman & Goldsmith, 2011). Studying expository texts on screen rather than on paper results in overconfident predictions of one’s performance on later tests. As a consequence, people do not devote enough study time to texts that are presented on screen and achieve low test performance.

**Metacognitive illusions of knowledge assessment**

Researchers have extensively examined the monitoring of whether information is available in memory – the feeling of knowing – with FOK judgments. FOK judgments are
subject to cue-familiarity illusions. In a study by Schwartz and Metcalfe (1992), participants studied word pairs for a cued recall test. For each target that participants could not recall at test, they indicated their confidence in identifying it in a later recognition memory test. Results showed higher FOK judgments when the pair’s cue word had been presented in an initial priming phase, even though priming did not affect actual memory. Reder and Ritter (1992) presented participants with arithmetic problems (e.g., $18 \times 27$) and asked them to quickly judge whether they could retrieve the answer from memory or had to compute it. People’s FOK judgments were the higher the more often they had seen parts of the problems in a previous training phase. This was even true when people did not have the answer in memory because the problems were novel (e.g., repeated exposure to $18 + 27$ increased FOK judgments for $18 \times 27$). These findings demonstrate that high familiarity of the information used to prompt FOK judgments can produce unduly high feelings of knowing that also fail to track differences in memory performance across conditions (or the lack thereof).

Another prominent illusion that afflicts FOK judgments involves high accessibility of invalid partial information. In a study by Koriat (1995), people answered general-knowledge questions and predicted their chances to recognize the correct answer in multiple-choice versions of the questions. Results showed that FOK judgments were based on the amount, intensity, and vividness of partial information that came to mind when searching for the answer in memory (e.g., someone expects that he knows the capital of Sweden because he can recall the capital of Denmark, watched a documentary about Norway, and feels knowledgeable about European geography). Reliance on accessibility of partial information produces illusory FOK judgments for deceptive items that bring to mind mostly incorrect information. For instance, the partial information people access when asked To what country does Corsica belong? typically points towards the wrong answer Italy rather than the correct answer France, resulting in too high FOK judgments. Thus, highly accessible invalid partial information produces unduly high feelings of knowing.

The uninformative-pictures illusion (Cardwell et al., 2017) demonstrates that the presence of irrelevant information can inflate people’s assessments of their knowledge. This illusion was found when people judged their knowledge of natural and mechanical processes. Cardwell et al. (2017) reported that judgments of understanding were higher when descriptive phrases such as how rainbows form or how radios work were preceded by uninformative photos (e.g., a picture of a rainbow in the sky). Presumably, the photos made it easier to generate relevant thoughts and images, which people erroneously interpreted as indicative of profound knowledge and understanding of the to-be-judged processes.

Metacognitive illusions of thinking

Metacognitive illusions also occur in thinking tasks. An example is the consistency illusion (Williams et al., 2020). Williams et al. (2020) presented six letters one of which was highlighted and instructed participants to find all five-letter words that began with the highlighted letter and could be formed by rearranging the presented letters. People’s retrospective confidence judgments about whether they had found all available words were higher when the highlighted letter was always in the same position (consistent) rather than when its position varied across trials (inconsistent). In contrast, position
consistency did not affect the number of words participants found. This finding shows that superficial procedural consistency can inflate confidence in one’s performance in thinking tasks.

Topolinski et al. (2016) reported the pronounceability illusion in quick judgments of solvability for anagrams (e.g., edisepo, solution: episode; ekisepo, no solution). People rated easy-to-pronounce anagrams (e.g., giloc, solution: logic) more often as solvable than hard-to-pronounce anagrams (e.g., ogili), even though easy-to-pronounce anagrams are harder to solve than hard-to-pronounce anagrams.

Text box 19.2 Creating a font-size illusion in class

This is a simplified and abbreviated version of Undorf and Zimdahl’s (2019) Experiment 2. Preprogrammed versions of this classroom demonstration (in English and German) including stimuli, instructions, and worksheets, as well as a template for analysis are available at https://osf.io/p5gau.

Method

Participants

The font-size illusion is quite large and should be found with 20 to 30 participants.

Materials

The experiment requires four font sizes. Select the smallest font size so that one can just read the words and select the largest font size so that the words fill a large part of the display. Then, select the two intermediate font sizes so that differences between all successive font sizes are equal perceptually. Undorf and Zimdahl (2019) used 9-point, 29-point, 93-point, and 294-point Arial fonts. The experiment requires 40 nouns of the same length (e.g., five letters). Assign ten randomly selected words to each font size.

Procedure

Ideally, all participants should view the stimuli from a similar distance. The experiment consists of three phases: a study phase, a filler task, and a test phase. Prior to the study phase, inform participants that they will study 40 words for a later test, in which they will be asked to write down as many study words as they can remember. Also inform participants that, during the study phase, they are required to estimate the chance of remembering each item at test on a percentage scale by writing down any number between 0 and 100. In the study phase, present each word for 2 seconds via projector. Immediately afterwards, present the JOL prompt “The chance to recall (0%–100%)?” Participants can write down their JOLs on a sheet of paper that lists word numbers in a column. After the study phase, participants work on an unrelated filler task for 90 seconds (e.g., solving arithmetic problems). In the test phase, participants are given 4 minutes to write down as many studied words as they can remember on a sheet of paper.
A font size illusion is present when JOLs increase with increasing font size, whereas recall performance is similar across all font sizes (see Figure 19.2 for typical results). A one-way ANOVA on JOLs using font size as a repeated-measures factor should reveal a significant effect. In contrast, a similar ANOVA on recall performance should be insignificant.

Theoretical accounts
Metacognitive illusions provide important insights into the basis of metacognition. First and foremost, systematic dissociations between metacognitive judgments and actual performance favor inferential accounts over direct-access accounts of metacognition (Dunlosky & Metcalfe, 2009; Koriat, 2007).

Direct-access versus inferential accounts of metacognition
Direct-access accounts of metacognition propose that metacognitive monitoring involves direct access to one’s own ongoing cognitive processes. According to direct-access accounts, JOLs rely on people’s observations of the memory traces that are formed during learning and of how trace strength increases as learning proceeds (Koriat, 1997). Similarly, FOK judgments are presumed to arise from searching one’s memory for traces of the currently unrecallable information and reading out whether a relevant trace exists and how strong it is (Hart, 1965). Because direct-access accounts assume that metacognitive judgments rely on the to-be-judged cognitions themselves, these accounts predict that metacognitive judgments are highly accurate throughout. Consequently, direct-access accounts of metacognition are incompatible with the metacognitive illusions we presented in the previous section and, more generally, with any systematic dissociation between metacognitive judgments and actual performance (e.g., Koriat, 2007).
Metacognitive illusions do, however, provide strong evidence for inferential accounts of metacognition. Inferential accounts assume that people infer the state of their cognitive system from cues and heuristics. JOLs are assumed to rely on cues that are available at study and pertain to the stimuli or learning conditions, such as semantic associations between the two words of a pair (e.g., no association: light – box; strong association: zebra – stripe; Koriat, 1997), the font size of study words (Rhodes & Castel, 2008), the length and complexity of sentences from to-be-studied texts (Maki et al., 2005), or whether texts are presented on screen or on paper (Ackerman & Goldsmith, 2011). FOK judgments are assumed to rely on the familiarity of the cues used to probe memory (Reder & Ritter, 1992; Schwartz & Metcalfe, 1992) and on the accessibility of partial information about the elusive information (Koriat, 1993, 1995).

According to inferential accounts of metacognition, metacognitive judgments are accurate when based on cues that are predictive of actual performance. In contrast, metacognitive illusions emerge when people base metacognitive judgments on invalid cues or fail to take cues into account that are predictive of actual performance. Examples of valid cues include “semantic association” or “text complexity” (Koriat, 1997; Maki et al., 2005). Basing one’s JOLs on the invalid cue “font size” produces the font-size illusion (see above and Text box 19.2), and people’s failure to incorporate the valid cue “future study opportunities” in their JOLs underlies the stability bias (see above).

Theory- and experience-based processes

Inferential accounts of metacognition distinguish between two types of processes underlying metacognitive judgments: theory-based and experience-based processes (e.g., Dunlosky & Tauber, 2014; Koriat & Levy-Sadot, 1999). Theory-based processes comprise the deliberate, analytic use of explicit beliefs and knowledge about cognition in general and about one’s own cognitive processes in particular. Prominent beliefs about memory include the beliefs that long-term memory has a limited capacity (held by 69% of a representative sample of 1,000 adult Norwegians; Magnussen et al., 2006), that dramatic events are remembered better than everyday events (70%; Magnussen et al., 2006), and that memory starts to decline by one’s 70th birthday at the latest (85%; Magnussen et al., 2006). When basing metacognitive judgments on theory-based processes, people are aware of the basis of their judgments (e.g., I expect to remember this situation, because it frightens me).

Experience-based processes are by-products of cognitive processes that influence people’s metacognitive judgments through non-analytic inferential processes operating below full awareness. Experience-based processes have the phenomenal quality of direct and unexplained intuitions (e.g., Koriat & Levy-Sadot, 1999). They often involve feelings of fluency, that is, the ease of processing information during reading, learning, retrieval, or problem solving. When basing metacognitive judgments on experience-based processes, people are unaware of the basis of their judgments (e.g., I will certainly remember this information, because I have a strong feeling that I will).

Theory-based and experience-based processes both contribute to accurate metacognitive judgments and metacognitive illusions. Correct explicit knowledge and beliefs may foster accurate metacognitive judgments. For instance, the belief that dramatic events are remembered better than everyday events may help learners to accurately predict that they will recall emotional stimuli better than neutral stimuli (Undorf et al., 2018; Zimmerman & Kelley, 2010). At the same time, basing metacognitive judgments on incomplete or faulty explicit beliefs about cognition harms metacognitive accuracy. For instance, people’s
knowledge about how emotion impacts memory is often incomplete insofar as they are unaware of the fact that memorial benefits of emotional stimuli do not extend to the cued recall of word pairs. Thus, basing JOLs on one’s belief about memory for emotional information may contribute to unduly high JOLs for emotional word pairs (Undorf & Bröder, 2020; Zimmerman & Kelley, 2010). Also, the incorrect belief that the capacity of long-term memory is limited may contribute to underestimating the exceptionally good memory performance for pictures and, consequently, underconfident JOLs (e.g., Undorf & Bröder, 2021).

It is important to note that holding metacognitive beliefs does not guarantee that these affect metacognitive judgments. For instance, the widely held belief that memory declines with old age did not affect the JOLs students made in seven experiments reported by Tauber et al. (2019). Often, metacognitive knowledge and beliefs must be activated to be incorporated into metacognitive judgments (Undorf & Erdfelder, 2015).

Basing metacognitive judgments on experience-based cues and heuristics that are predictive of actual performance results in accurate metacognitive judgments, whereas reliance on invalid experience-based cues and heuristics produces metacognitive illusions. Because the validity of non-analytic cues can vary between situations, reliance on one and the same cue may produce accurate metacognitive judgments under some conditions and metacognitive illusions under other conditions. One example is retrieval fluency. Koriat and Ma’ayan (2005) found that basing JOLs for word pairs on the ease of retrieving information from memory contributed to JOL accuracy. In their study, JOLs were elicited with a delay after studying. Prior to making each JOL, participants saw the first word of the respective pair and were asked to recall the second word. The quicker participants retrieved the second word prior to making their JOLs, the higher were their JOLs and their performance on a final test. Thus, basing JOLs on retrieval fluency contributed to accurate JOLs. At the same time, however, retrieval fluency underlies the semantic-fluency illusion, which is due to basing JOLs for the free recall of words on the fluency of retrieving these words in a completely unrelated task (see above).

In recent years, many studies addressed the relative contributions of experience-based and theory-based processes to metacognitive illusions (for a review, see Undorf, 2020). This literature suggests, for instance, that the font-size illusion is mainly due to explicit beliefs about memory (but see Yang et al., 2018). People were found to believe that words printed in larger fonts are easier to remember than words printed in smaller fonts (e.g., Mueller et al., 2014; Undorf & Zimdahl, 2019). Also, independent measures of fluency such as self-paced study time or lexical decision time were similar for small-font and large-font words (e.g., Mueller et al., 2014). Most importantly, using a wide range of font sizes (6 point to 500 point) dissociated perceptual fluency and JOLs. While JOLs increased with font size across the whole range of font size, perceptual fluency was lower for very small and very large font sizes than for intermediate font sizes (e.g., Undorf & Zimdahl, 2019).

In contrast, other illusions were found to be due to experience-based processes. One example is the stimulus-size illusion (Undorf et al., 2017). This illusion affects JOLs for visual stimuli that gradually increase in size: Stimuli that increased quickly rather than slowly received higher JOLs, even though clarification speed did not affect memory performance. The finding that quick clarification did not increase JOL directly, but only indirectly through higher fluency (independently measured as the time it took people to identify the stimuli) indicates that the stimulus-size illusion is due to experience-based...
Metacognitive illusions

processes. Moreover, participants were unaware of differences in clarification speed and did not believe quick clarification to help memory (Undorf et al., 2017).

In summary, the very existence of metacognitive illusions demonstrates that metacognitive judgments are inferences from cues people regard as informative about cognition in general and their own cognition in particular. Metacognitive illusions emerge when people base their metacognitive judgments on invalid cues or fail to rely on valid cues. The cues that produce metacognitive illusions can be broadly categorized into (1) faulty beliefs about cognition that affect metacognitive judgments through theory-based processes and (2) invalid non-analytic cues and heuristics that affect metacognitive judgments through experience-based processes.

Applied perspectives

Metacognitive monitoring plays an integral role for effective self-regulation of cognition and behavior (e.g., Bjork et al., 2013; Metcalfe, 2009). Illusions and biases in metacognitive judgments can therefore lead to ineffective self-regulation and poor performance in cognitive tasks.

Real-life consequences of metacognitive illusions

There is ample evidence for detrimental effects of metacognitive illusions and biases on cognitive performance (cf. Text box 19.1 above). Many studies addressed the negative consequences of overconfidence (see Chapter 18). For instance, Dunlosky and Rawson (2012) provided clear evidence that overconfidence can produce underachievement. In their experiments, students learned key-term definitions across alternating study and test phases (e.g., The just-world hypothesis is the strong desire or need people have to believe that the world is an orderly, predictable, and just place, where people get what they deserve). In each test phase, students judged the accuracy of their recalls. Individual differences in overconfidence predicted performance on the final test: The less overconfident people were when judging the accuracy of their recalls, the better was their final test performance. Also, experimentally reducing overconfidence in the students’ judgments improved their final test performance.

Other research focused on adverse effects of specific metacognitive illusions. An educationally relevant illusion is the spacing (or interleaving) illusion, that is, people’s failure to appreciate that long-term retention and inductive reasoning benefit from the spacing, rather than massing, of practice trials (e.g., Kornell & Bjork, 2008a; Logan et al., 2012). This illusion contributes to learners’ robust preference for blocking repeated study trials (Tauber et al., 2013) and for restudying high priority information sooner rather than later (Cohen et al., 2013), even though restudying after (longer) intervals is more effective.

The surprising finding that learners who have the option to self-regulate their learning by dropping flashcards perform less well than learners who may not drop flashcards and, consequently, spend similar study time on all information, has been connected to the stability bias mentioned above (Kornell & Bjork, 2008b). Presumably, learners who underestimate the learning that might result from additional studying and disregard future forgetting drop flashcards too early and, consequently, perform poorly.

Although often studied from an educational perspective, negative effects of metacognitive illusions on cognitive performance are by no means restricted to educational settings. For instance, Hargis and Castel (2018) discuss adverse consequences of metacognitive
illusions on health. The authors focus on how metacognitive illusions may contribute to the difficulty of taking medications as prescribed, which is experienced by many people and older adults in particular. Overconfidence in one’s own learning and memory of which medication to take in what dosage or at what time of day may prevent patients from taking notes in the doctor’s office or from recognizing that they have forgotten important medication-related information. The stability bias may contribute to patients overestimating how far into the future they will remember medication-related information provided by the doctor. This example demonstrates that metacognitive illusions play a role in the healthcare sector, with potentially dramatic implications for patients and considerable costs for society (e.g., falls, heart failure, and preventable hospitalizations).

**Mending metacognitive illusions**

In view of the adverse consequences of inaccurate metacognitive judgments, an obvious question is whether metacognitive illusions can be mended. Research suggests that this is indeed possible, even though it is not necessarily easy. Sidi et al. (2017) demonstrated that alleviating screen inferiority is relatively easy: Emphasizing that tasks were of high importance eliminated people’s overconfidence in their solutions to challenging problems when the problems were presented on the screen rather than on paper. Similarly, Koriat et al. (2004) found that JOLs were sensitive to future forgetting when the announced retention interval was manipulated within participants rather than between participants: JOLs declined monotonically with retention interval when participants knew during study that some items would be tested after 10 minutes, whereas other items would be tested after one day or one week.

In other cases, however, metacognitive illusions are difficult to overcome. One example is the interleaving illusion. In a study by Yan et al. (2016), people learned the styles of artists from examples of the artists’ paintings. During study, paintings were presented blocked by artist for some artists and interleaved for other artists. On the test, participants assigned new paintings to the studied artists. Almost all participants achieved better classification performance for artists whose paintings were studied interleaved rather than blocked. After the test, however, they incorrectly judged that blocking was more effective than interleaving. Even when receiving personalized feedback about their test performance and detailed explanations why interleaving is superior to blocking, at least 50% of participants continued to show the interleaving illusion. After conducting six experiments with a total of more than 700 participants, Yan et al. (2016) concluded that “mending the metacognitive illusion that blocking is more effective for learning proved a daunting task” (p. 931). Why is the interleaving illusion so difficult to overcome? Yan et al. (2016) identified three reasons. First, learners experience blocked practice as easier and more fluent than interleaved practice. Second, most people have a strong belief that blocking is more effective than interleaving. Finally, learners often assume that general principles of learning may not apply to themselves. When informed that interleaving leads to better learning in 90% of learners, between 48% and 69% of the learners placed themselves in the remaining 10% of learners.

Mending metacognitive illusion is further complicated by the fact that successful debiasing may not generalize beyond the context in which it originally took place (e.g., Koriat & Bjork, 2006). Koriat and Bjork (2006) examined procedures for alleviating the foresight bias, that is, inflated JOLs for to-be-remembered information that appears obvious at study but is relatively difficult to remember at test (e.g., water – well appears semantically
related and highly memorable at study, but when water is presented alone at test, many other likely responses such as drink, cool, and swim come to mind and reduce memory performance. Two different debiasing procedures effectively alleviated the foresight bias. First, studying the same items across multiple study-test cycles considerably reduced the foresight bias beginning with the second study-test cycle. Second, educating participants about the foresight bias after the first study-test cycle and instructing them to avoid this bias resulted in a similar amount of debiasing from the second study-test cycle on. Importantly, however, only the debiasing achieved by explicit instruction yielded transfer to a new set of items (Koriat & Bjork, 2006).

Despite these difficulties in mending specific metacognitive illusions, it is certainly possible to create conditions that promote high monitoring accuracy. For instance, metacognitive monitoring accuracy in the classroom can be improved by using the wait-generate-validate strategy (Hausman et al., 2021). Wait refers to the recommendation that one should judge one’s understanding not immediately after learning but should wait for a couple of hours or one day. After this delay, one should try to actively generate the material from memory without the help of study materials. Finally, one needs to validate the accuracy and completeness of the information one has generated. When following these steps, it is relatively easy to accurately assess one’s learning and understanding. The resulting improvement in monitoring can significantly increase learning.

In summary, debiasing and creating favorable conditions for accurate metacognitive monitoring are promising approaches to ameliorating the negative effects of metacognitive illusions on cognitive performance.

Conclusions

Metacognitive illusions have attracted a great deal of interest from researchers and, because of their detrimental effects on cognitive performance, from practitioners. There is no doubt that the study of metacognitive illusions has substantially advanced our understanding of metacognition. However, it is important to keep in mind that metacognitive illusions are by-products of functional metacognitive processes and that metacognition is quite accurate by and large.

Summary

• Metacognition refers to people’s assessments of their cognition (metacognitive monitoring) and to the use of this information for self-regulation (metacognitive control).
• Metacognitive illusions are systematic dissociations between cognition and metacognition that have been found in domains such as learning, knowledge assessment, and thinking.
• Metacognitive illusions undermine self-regulation and performance in cognitive tasks.
• The existence of metacognitive illusions indicates that people cannot directly access their cognition, but infer the state of their cognitive system from cues and heuristics.
• Faulty explicit beliefs about cognition as well as invalid non-analytic cues and heuristics contribute to metacognitive illusions.
• The negative effects of metacognitive illusions on cognitive performance can be ameliorated by debiasing techniques and by creating favorable conditions for assessing one’s cognitive processes.
Further reading


Acknowledgment

Manuscript preparation was supported by grants UN 345/1-3 and UN 345/2-1 from the Deutsche Forschungsgemeinschaft and by a Margarete von Wrangell fellowship from the state of Baden-Württemberg to Monika Undorf.

References


Metacognitive illusions


