



# The many facets of metacognition: comparing multiple measures of metacognition in healthy individuals

Anneke Terneusen<sup>1,2</sup> · Conny Quaedflieg<sup>1</sup> · Caroline van Heugten<sup>1,2,3</sup> · Rudolf Ponds<sup>3,4</sup> · Ieke Winkens<sup>1,2</sup>

Received: 21 December 2022 / Accepted: 10 June 2023 / Published online: 7 July 2023  
© The Author(s) 2023

## Abstract

Metacognition is important for successful goal-directed behavior. It consists of two main elements: metacognitive knowledge and online awareness. Online awareness consists of monitoring and self-regulation. Metacognitive sensitivity is the extent to which someone can accurately distinguish their own correct from incorrect responses and is an important aspect of monitoring of behavior. Research into the interplay between these elements is currently lacking. Therefore, the aim of the current study was to explore how these different elements of metacognition can predict metacognitive sensitivity. Healthy participants filled out the Metacognitive Awareness Inventory with two subscales that serve as measures of metacognitive knowledge. Next, as measures of online awareness, they performed a memory task that was adapted to include trial-by-trial confidence judgments, and made pro- and retrospective confidence judgments about their performance on this task. The 128 included participants show a large variability in scores on all the different metacognitive measures. Metacognitive sensitivity was predicted by higher prospective discrepancy scores and lower retrospective discrepancy scores but not by metacognitive knowledge. The current study confirms that metacognition is a multidimensional concept consisting of different elements. Online measures seem to be associated with each other, but not so much with offline measures. The current framework can be used to further investigate the associations between different elements of metacognition within persons.

**Keywords** Metacognition · Metacognitive sensitivity · Metacognitive knowledge · Prospective confidence judgments · Retrospective confidence judgments

---

✉ Anneke Terneusen  
anneke.terneusen@maastrichtuniversity.nl

<sup>1</sup> Department of Neuropsychology and Psychopharmacology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>2</sup> Limburg Brain Injury Centre, Maastricht University, Maastricht, The Netherlands

<sup>3</sup> Department of Psychiatry and Neuropsychology, MHeNS School for Mental Health and Neuroscience, Maastricht University, Maastricht, The Netherlands

<sup>4</sup> Department of Medical Psychology, Amsterdam University Medical Center, location VUmc, Amsterdam, Netherlands

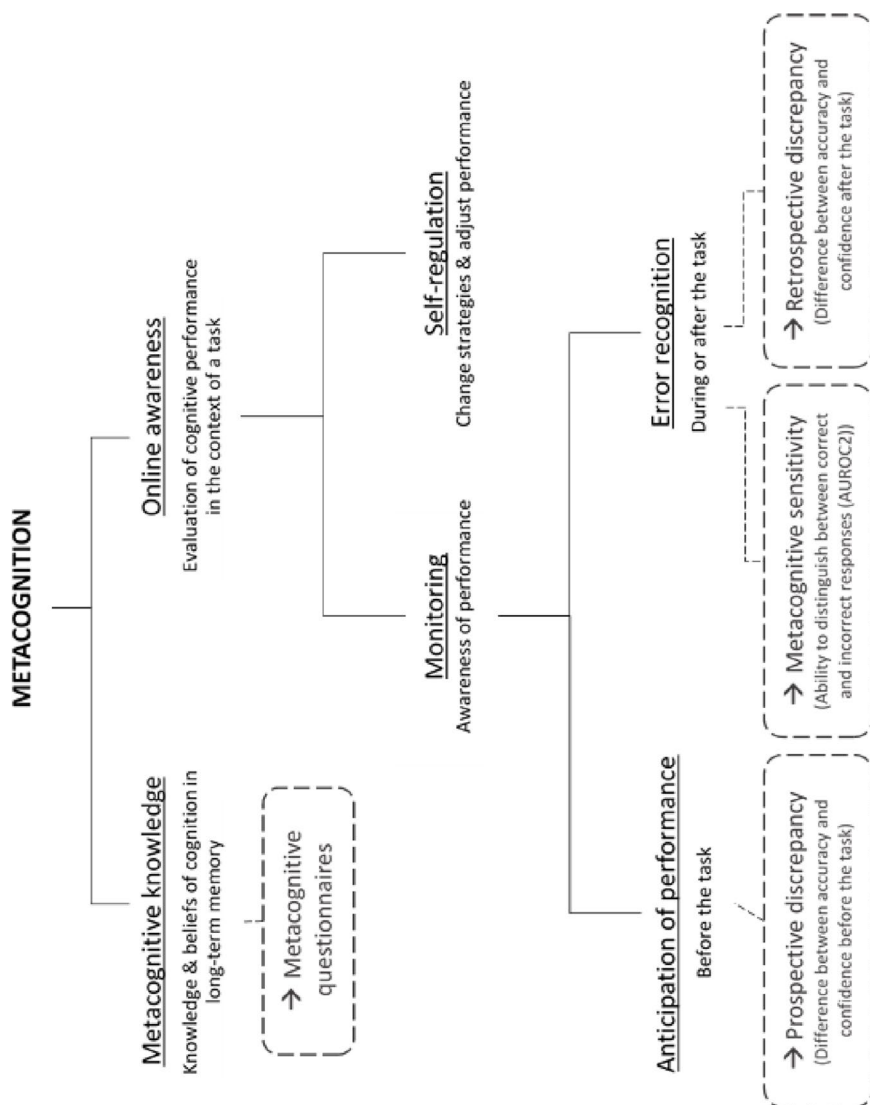
## Introduction

Metacognition is the ability to be conscious of your own thoughts, actions, and cognitions (Flavell, 1979; Stuss, 1991). This monitoring is necessary to adapt and achieve successful goal-directed behavior. There are many ways to define metacognition. Most models are based on a distinction between a stable reflective knowledge of one's cognitive capacities and a more "on the fly" online metacognition (Fleur et al., 2021). Another distinction can be made between local metacognition, such as confidence in isolated decisions, and global metacognition such as self-efficacy beliefs (Seow et al., 2021). Fleur et al. (2021) stress the need for more studies on protocols to measure the different constructs of metacognition and the relationships between these constructs.

Two main elements of metacognition, as described by a neuropsychological model of metacognition, are metacognitive knowledge and online awareness (Fig. 1; Toglia & Kirk, 2000). Metacognitive knowledge is knowledge and beliefs of one's cognitive processes and strategies that are stored in long-term memory. This knowledge base is gained over time and is rather stable, although it can be influenced by successes and failures (Toglia & Kirk, 2000). Online awareness refers to evaluation of performance within the context of a task. This is more flexible as it depends on task characteristics such as complexity and familiarity of the task, as well as personal factors such as motivation and meaningfulness of the task (Toglia & Kirk, 2000). This resembles "on the fly" metacognition (Fleur et al., 2021) or metacognitive experiences (Efklides, 2009). In this model, online awareness consists of two elements. The first element is monitoring, which is awareness of performance within the context of a task. This includes anticipating performance after appraisal of task demands but before task completion, as well as error recognition during a task. The second element is self-regulation, which refers to the ability to change strategies and adjust performance in response to changing task demands and experience. These self-regulatory processes depend on accurate self-monitoring but also motivation and socio-emotional processes. There are different measurement methods to assess these different elements of metacognition (see Fig. 1).

Common ways of assessing metacognition are questionnaires and performance-based measures. Metacognitive questionnaires typically address situations either in the past or in the future, which trigger pre-existing knowledge and beliefs about one's cognition stored in long-term memory and, therefore, quantifies metacognitive knowledge. Online awareness can be quantified with performance-based measures. One way is using discrepancy scores between performance accuracy and confidence in one's performance assessed before and after performing the task (Fleming et al., 2016). Confidence judgments before a task (prospective confidence) are based on an analytic process of evaluating task demands and experiences (Koriat, 2000), for which one might employ metacognitive knowledge. Confidence judgments made after a task (retrospective confidence) are experience-based judgments that rely on feeling of knowing (Koriat, 2000) and is an overall judgment of performance (e.g. this was a difficult task for me, I think I have not done well).

There are also more computational measures of metacognition and these have developed over time. For example, the Goodman-Kruskal gamma correlation was one of the first recommended measures for accuracy of feeling-of-knowing (Nelson, 1984). Others have investigated metacognition using signal detection theory (Benjamin & Diaz, 2008) or have investigated relative and absolute accuracy of feeling-of-knowing judgments (Schwartz et al., 2016). Furthermore, there are non-parametric methods such as metacognitive sensitivity (Fleming & Lau, 2014). Metacognitive sensitivity can be measured by



**Fig. 1** Conceptual overview of metacognition based on previous literature. *Note.* This figure provides a conceptual overview of metacognition. The arrows are the measurement methods to assess the different elements of metacognition, which are underlined. AUROC2 = type two area under the receiver operating characteristic curve

asking participants during the task, after each trial, to indicate how confident they are that their answer is correct. When someone reports high confidence for correct answers and low confidence for incorrect answers, they are considered to have good metacognitive sensitivity (Fleming & Lau, 2014).

The assessment of metacognition remains difficult and there is a demand for empirical evidence to support theoretical frameworks of metacognition and its different elements (Seow et al., 2021). Moreover, there is a gap in the literature concerning the associations between the different constructs of metacognition (Fleur et al., 2021). Therefore, the aim of the current study was to explore how metacognitive sensitivity can be predicted by the different elements of metacognition, as measured by different measurement methods. The hypotheses are that smaller prospective and retrospective discrepancy scores (i.e. better metacognition) will predict better metacognitive sensitivity. In addition, metacognitive knowledge is expected to be positively associated with metacognitive sensitivity.

## Methods

### Participants

Participants were 128 healthy volunteers (100 female, 27 male, 1 non-binary) who were 20.6 ( $\pm 1.6$ ) years old. Most were not native English speakers ( $N=116$ ). Participants had to (1) be 19 up to and including 24 years old, (2) be able to give informed consent, (3) have a good comprehension of the English language. Participants were excluded when they (1) had a history of, or a current, psychiatric illness or neurodegenerative disease, or (2) had a history of, or were currently under treatment for, alcohol or substance abuse. Participants were recruited through flyers at the Maastricht University and social media. In order to reach a power of 0.80, with a significance level of 0.05, and a medium effect size of  $f^2=0.10$ , the aim was to include a minimum of 124 participants. The final sample consisted of 128 participants (37 exclusions due to invalid answers).

### Procedure

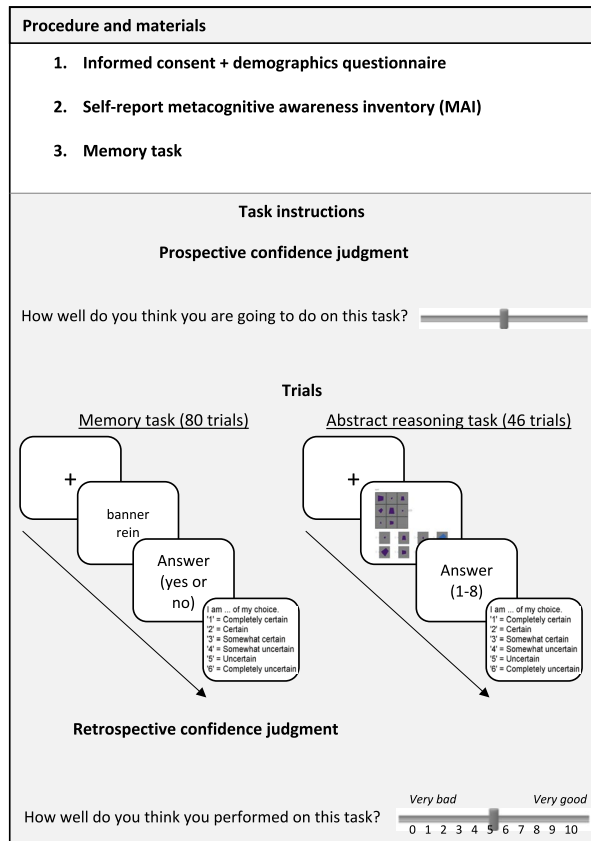
Participants completed the study from home in a single online session that took about an hour using Qualtrics (see Fig. 2). Participants first read the information letter and signed informed consent online. After filling-out a short demographics questionnaire, the self-report metacognitive questionnaire was filled out. Subsequently, they performed a memory task and an abstract reasoning task (see Supplementary materials). Some validity questions were entered throughout the session to ensure that participants took part in a serious manner. Ethical approval was obtained from the Ethics Review Committee Psychology and Neuroscience at Maastricht University with reference number OZL\_233\_21\_02\_2021.

## Materials

### Metacognitive Awareness Inventory (MAI)

This self-report questionnaire developed by (Schraw & Dennison, 1994) consists of 52 items measuring metacognitive knowledge. Ratings for each item were made on a

**Fig. 2** Schematic overview of study procedure and materials. *Note.* MAI = Metacognitive Awareness Inventory



continuous scale ranging from 0 (completely false) to 100 (completely true). Psychometric analyses of this instrument show that it is a reliable instrument with high internal consistency (coefficient  $\alpha = 0.88$  per subscale; Schraw and Dennison, 1994).

### Prospective and retrospective confidence judgments

For the tasks, prospective confidence judgments about performance (after reading the instructions but before starting the task) and retrospective confidence judgments about performance (directly after completion of the task) were made. Participants had to indicate how well they thought they would perform (prospective) or had performed (retrospective) on the task on a scale from 0 to 10, where 0 corresponded to ‘very bad’ and 10 to ‘very good’.

### Memory task

The memory task was a word pair learning task consisting of two phases: an encoding phase and a recognition phase (Inquisit 6, (Millisecond, 2006). During the encoding phase, participants were presented with 40 unrelated word pairs (cue word + pair) which were previously used and validated by Payne et al. (2012; List 3). Each word pair was displayed for 5 seconds and order of presentation was randomized for each participant. The recognition

phase consisted of a two-alternative forced choice test. Participants were presented with 40 word pairs again. Cue words from the learned word pairs list were either paired with a correct cue word (in 50% of trials), or a new distractor word. They had to determine whether the word pair was in the list they had previously learned. Answer options were ‘yes’ or ‘no’. The order of presentation was random, there was no time limit to give an answer, and participants were asked to be as accurate as possible. Every trial each answer was followed by a confidence rating. Participants had to indicate how confident they were their answer was correct. There were six options ranging from ‘completely uncertain’ to ‘completely certain’. Participants were instructed that ‘completely uncertain’ meant that they answered completely based on guessing.

## Outcome measures

### Error recognition

**Metacognitive sensitivity** Whether a person can differentiate correct from incorrect responses answer was measured using the trial-by-trial confidence ratings and accuracy of the answers given. Metacognitive sensitivity was quantified with the type two area under the receiver operating characteristic curve (AUROC2) code from Fleming and Lau (2014) in MATLAB (R2021b; Inc, 1996). Scores can range from 0 to 1. A score of 0.5 is at chance and corresponds to no metacognitive sensitivity. A score above 0.5 indicates higher metacognitive sensitivity, with 1 representing perfect metacognitive sensitivity (people are high confident for correct trials, and low confident for incorrect trials). Scores below 0.5 are atypical but indicate reversed metacognitive sensitivity (people are high confident for incorrect trials, and low confident for correct trials).

**Retrospective discrepancy** Discrepancy scores between the confidence judgments made after the tasks (ranging from 0 to 10) and performance accuracy (number of trials correct/total number of trials\*10; ranging from 0 to 10) were calculated. The discrepancy scores can range from –10 up to 10. Scores close to zero represent good metacognition and the further the score is away from zero (either negative or positive), the more this type of metacognition is impaired.

### Metacognitive knowledge

**Metacognitive Awareness Inventory (MAI)** The *knowledge of cognition* subscale of the self-report questionnaire MAI consists of questions covering knowledge about self and about strategies, knowledge about how to use strategies, and knowledge about when and why to use strategies (Schraw & Dennison, 1994). Mean scores range from 0 (low) to 100 (high) level of perceived knowledge of cognition.

The *regulation of cognition* subscale of the self-report questionnaire MAI consists of questions covering the control aspect of learning: planning, information management strategies, comprehension monitoring, debugging strategies, and evaluation strategies (Schraw & Dennison, 1994). Mean scores range from 0 (low) to 100 (high) level of perceived regulation of cognition.

## Anticipation of performance

**Prospective discrepancy** Discrepancy scores between the confidence judgments made before the tasks (ranging from 0 to 10) and performance accuracy (number of trials correct/total number of trials\*10; ranging from 0 to 10) were calculated. The discrepancy scores can range from -10 up to 10. Scores close to zero represent good metacognition and the further the score is away from zero (either negative or positive), the more this type of metacognition is impaired. Negative scores represent under-confidence and positive scores represent over-confidence.

## Data analysis

Statistical analyses were conducted with IBM SPSS Statistics version 27 for Windows. All measures showed high reliability scores. Cronbach's alpha was  $\alpha=0.827$  on the MAI knowledge of cognition subscale and  $\alpha=0.859$  on the regulation of cognition subscale. Odd-even split-half correlation analyses showed high reliability on the memory task ( $r=.828$  for accuracy;  $r=.935$  for confidence ratings). Descriptive analyses were performed to identify the distribution of scores on the different metacognitive measures (see Supplementary materials). Multiple regression analyses were conducted to investigate whether scores on the questionnaires and accuracy of confidence judgments could predict metacognitive sensitivity (AUROC2). For these analyses, the prospective and retrospective discrepancy scores were transformed into absolute values (because both positive and negative values indicate poor metacognition). The raw data is available at <https://doi.org/10.34894/U1YTW8>. The experiment was not preregistered.

## Results

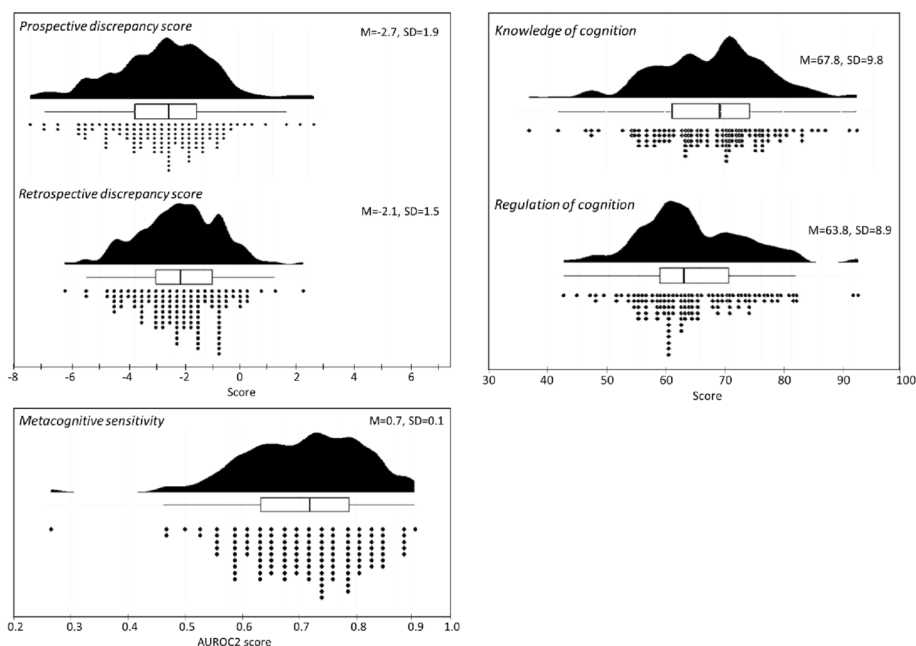
The distribution of scores on the metacognitive measures for the memory task can be found in Fig. 3. Performance scores can be found in Table 1. Performance accuracy on the abstract reasoning task was close to guessing ( $M=6.1$ ,  $SD=2.1$ ). Therefore, the metacognitive measures on this task cannot be reliably interpreted and will not be further discussed (but can be found in Supplementary materials).

### Prospective and retrospective discrepancy scores predict metacognitive sensitivity

The multiple regression model significantly predicted metacognitive sensitivity ( $F_{(4,123)}=2.637$ ,  $p<.05$ ,  $R^2=0.079$ , see Table 2). Significant predictors were higher prospective discrepancy scores ( $\beta=0.233$ ,  $p=.014$ ) and lower retrospective discrepancy scores ( $\beta=-0.185$ ,  $p=.038$ ). MAI knowledge of cognition and regulation of cognition subscales did not significantly predict metacognitive sensitivity (all  $p>.703$ ).

## Discussion

The aim of this study was to explore whether metacognitive knowledge, anticipation and evaluation of performance predict metacognitive sensitivity. In this healthy population there is large variability of scores on all the different metacognitive measures. The results show that metacognitive sensitivity on the memory task was predicted by poor anticipation of performance (large prospective discrepancy score) and good retrospective evaluation of performance (small retrospective discrepancy score). Moreover, metacognitive knowledge did not predict metacognitive sensitivity.



**Fig. 3** Distribution of scores on the different metacognitive measures. *Note.* The raincloud plots visualize the raw data and show the distribution of the dataset on the memory task. The boxplots show medians and quartiles. The dots are the raw data points. AUROC2=type two area under the receiver operating characteristic curve

The hypothesis that there would be a negative association between metacognitive sensitivity and discrepancy scores was true for the retrospective discrepancy score but not the prospective discrepancy score. Higher metacognitive sensitivity during the task was associated with more accurate judgment of performance after the task. This suggests that people use the metacognitive information during the task to shape their evaluation after the task, as has been suggested in previous research (Akama & Yamauchi, 2004). The confidence judgment after the task is different from before the task, and this might be due to metacognitive experiences during the task such as familiarity, difficulty, satisfaction, and effort (Efklides, 2006). The shift from a positive association between metacognitive sensitivity and prospective discrepancy to a negative association with retrospective discrepancy supports this idea. Though the proportion of variance was small, indicating that there are likely other important predictors for metacognitive sensitivity, such as domain-specific self-concept (Händel et al., 2020).

While online metacognitive measures seem to be associated with each other, offline measures seem not to. Metacognitive knowledge did not predict metacognitive sensitivity. This confirms

**Table 1** Accuracy scores on the memory task

<i>N</i> = 128	Mean ( $\pm$ SD)	Min – max
Memory task		
Accuracy (out of 10)	8.2 ( $\pm$ 1.2)	3.8–10
Prospective confidence rating (out of 10)	5.6 ( $\pm$ 1.6)	0.0–9.0
Retrospective confidence rating (out of 10)	6.1 ( $\pm$ 2.0)	1.0–10.0
Mean confidence on trial-by-trial ratings (out of 6)	4.4 ( $\pm$ 0.8)	2.4–6.0



**Table 2** Multiple regression analysis predicting metacognitive sensitivity

Dependent variable	Independent variables	R <sup>2</sup>	Adjusted R <sup>2</sup>	Standardized beta	p-value
Memory task Metacognitive sensitivity (AUROC2)	Knowledge of cognition (MAI)	0.079	0.049	0.046	0.703
	Regulation of cognition (MAI)			<0.001	1.000
	Prospective discrepancy score (absolute values)			0.223	<b>0.014</b>
	Retrospective discrepancy score (absolute values)			−0.185	<b>0.038</b>

AUROC2 = type two area under the receiver operating characteristic curve. MAI = Metacognitive Awareness Inventory

the idea that metacognition can be split into metacognitive knowledge and online awareness as described in different models (e.g. Flavell, 1979; Toglia & Kirk, 2000). The questionnaire items cover a broad range of metacognitive behaviors and strategies, they relate to situations more distant in time, and might be more abstract than the in-the-moment confidence judgments after every trial. This is in line with a recent review that concluded that self-report measures can be unclear and differ from metacognitive behavior (Craig et al., 2020).

The strengths of this study are the large sample size, and the inclusion of different measures of metacognition, which allows for a proper investigation of the associations between the different elements of metacognition. However, there are also some limitations to this study. The first is that the study was fully online, making it difficult to check whether people took part seriously or whether they took long breaks. When replicating this study in a more controlled environment such as a lab, this can be registered. For now, we attempted to control for this by adding validity questions. These questions led to exclusion of five participants who indicated they ‘just clicked through’ the study. Secondly, the included sample was very homogenous, including mostly highly educated healthy people within a small age range. This limits generalizability of the results. Thirdly, the metacognitive sensitivity measure (AUROC2) might not be very meaningful when there is a floor or ceiling effect in performance accuracy. There must be enough correct and incorrect trials for people to be able to distinguish the two types of trials with their confidence judgment. Similarly, if the task is too difficult and people do not know whether their answer is correct or incorrect, it is difficult to give a confidence judgment. Because performance was at chance level, the abstract reasoning task was disregarded in further interpretation. There is not enough metacognitive information to be gained during the task. In future research it is important to adopt a task that is not too easy because people need to have both correct and incorrect trials, but also not too difficult because people need to know to some degree whether their answer was correct or not. A solution could be to use an adaptive task. Furthermore, in replication studies, if multiple tasks are included, they should be counterbalanced or randomized to control for practice effects or fatigue.

The current study shows that task-specific online measures of metacognition can predict metacognitive sensitivity, but more general self-report measures of metacognition cannot. Moreover, it provides a framework to investigate the associations between different elements of metacognition within persons. An interesting application of this paradigm would be to investigate these associations in populations with metacognitive impairments, such as after brain disorders. This could give an indication of where the problem arises and, possibly, what should be targeted to improve metacognition. Moreover, for a more comprehensive understanding of metacognition, it would be interesting to investigate which brain areas or networks are involved in the different elements of metacognition.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s11409-023-09350-1>.

**Acknowledgements** Authors would like to thank Sheila Kirchgessner for her help with data collection.

**Data Availability** The raw data is available at <https://doi.org/10.34894/U1YTW8>.

## Declarations

**Ethical approval** For this online correlational study involving human participants, ethical approval was obtained from the Ethics Review Committee Psychology and Neuroscience at Maastricht University with reference number OZL\_233\_21\_02\_2021.

**Conflict of interest** None.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Akama, K., & Yamauchi, H. (2004). Task performance and metacognitive experiences in problem-solving. *Psychological Reports*, 94(2), 715–722.
- Benjamin, A. S., & Diaz, M. (2008). Measurement of relative metamnemonic accuracy. In Dunlosky, J., & Bjork, R. A. (Eds.), *Handbook of Metamemory and Memory* (pp. 73–94). New York, NY: Psychology Press.
- Craig, K., Hale, D., Grainger, C., & Stewart, M. E. (2020). Evaluating metacognitive self-reports: Systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning*, 15, 155–213.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3–14.
- Efklides, A. (2009). The role of metacognitive experiences in the learning process. *Psicothema*, 21(1), 76–82.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring- A new area of cognitive-developmental inquiry. *The American Psychologist*, 34(10), 906–911.
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, 443. <https://doi.org/10.3389/fnhum.2014.00443>
- Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J. C. (2016). Metacognition about the past and future: Quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, 2016(1), niw018. <https://doi.org/10.1093/nc/niw018>
- Fleur, D. S., Bredeweg, B., & van den Bos, W. (2021). Metacognition: Ideas and insights from neuro-and educational sciences. *NPJ Science of Learning*, 6(1), 1–11.
- Händel, M., de Bruin, A. B. H., & Dresel, M. (2020). Individual differences in local and global metacognitive judgments. *Metacognition and Learning*, 15(1), 51–75. <https://doi.org/10.1007/s11409-020-09220-0>
- Inc., T. M. (1996). *MATLAB*. In (Version R2021b).
- Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, 9(2), 149–171.
- Millisecond, I. (2006). (Version 6).
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95(1), 109.
- Payne, J. D., Tucker, M. A., Ellenbogen, J. M., Wamsley, E. J., Walker, M. P., Schacter, D. L., & Stickgold, R. (2012). Memory for semantically related and unrelated declarative information: The benefit of sleep, the cost of wake. *PLoS One*, 7(3), e33079. <https://doi.org/10.1371/journal.pone.0033079>
- Schraw, G., & Dennison, R. S. (1994). Metacognitive awareness inventory. *Contemporary Educational Psychology*, 19, 460–475.
- Schwartz, B. L., Boduroglu, A., & Tekcan, A. (2016). Methodological concerns: The feeling-of-knowing task affects resolution. *Metacognition and Learning*, 11, 305–316.
- Seow, T. X., Rouault, M., Gillan, C. M., & Fleming, S. M. (2021). How local and global metacognition shape mental health. *Biological Psychiatry*, 90(7), 436–446.
- Stuss, D. T. (1991). Self, awareness, and the frontal lobes: A neuropsychological perspective. *The self: Interdisciplinary approaches* (pp. 255–278). Springer.
- Toglia, J., & Kirk, U. (2000). Understanding awareness deficits following brain injury. *NeuroRehabilitation*, 15, 57–70.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.