



OPEN ACCESS

EDITED BY

Christopher A. Was,
Kent State University,
United States

REVIEWED BY

John Dunlosky,
Kent State University,
United States
Silvia Chapman,
Columbia University Irving Medical Center,
United States

*CORRESPONDENCE

Michael K. Yeung
mkcyung123@gmail.com

SPECIALTY SECTION

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

RECEIVED 11 April 2022

ACCEPTED 04 August 2022

PUBLISHED 22 August 2022

CITATION

Yeung MK (2022) Stability and consistency
of metamemory judgments within a
session.

Front. Psychol. 13:917576.

doi: 10.3389/fpsyg.2022.917576

COPYRIGHT

© 2022 Yeung. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Stability and consistency of metamemory judgments within a session

Michael K. Yeung^{1,2*}

¹Department of Rehabilitation Sciences, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China, ²University Research Facility in Behavioral and Systems Neuroscience, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR, China

There has been much interest in assessing individual and group differences in metamemory skills. These endeavors require or would benefit from enhanced knowledge about the stability and consistency of metamemory judgments over successive trials. However, few studies have examined these aspects. Thus, the present study investigated and compared the within-session stability and consistency of three major types of metamemory judgment: judgment of learning (JOL), feeling-of-knowing judgment (FOK), and retrospective confidence judgment (RCJ), using a single-task paradigm. A total of 38 healthy young adults (17 males, 21 females) completed three trials of a face–scene associative learning task designed to assess metamemory. In each trial, participants rated their JOLs while studying a new set of face–scene pairs, and then rated their FOKs and RCJs while their memory was being tested. The stability and consistency of the mean confidence ratings and the relationships between confidence rating and memory performance, indexed by two gamma estimates, were analyzed and compared across types of metamemory judgments. Over trials, there was a significant decrease in the mean rating for JOL but not for FOK or RCJ. Also, the gamma scores of JOL, but not that of FOK or RCJ, significantly improved with practice. Furthermore, for each type of metamemory judgment, the mean confidence rating showed excellent consistency across trials. Depending on the judgment type and gamma estimation method, the consistency of gamma scores ranged from poor to excellent. Thus, the present study clarified the temporal dynamics of various types of metamemory judgments and the consistency of metamemory measures.

KEYWORDS

judgment of learning, feeling-of-knowing, retrospective confidence judgments, memory, metacognition, gamma

Introduction

Metamemory refers to knowledge about one's own memory ability, and it involves the interaction between monitoring and control processes (Nelson and Narens, 1990; Koriat and Goldsmith, 1996; Metcalfe, 2000). Metamemory monitoring is believed to play a pivotal role in influencing learning behavior. For example, accurate knowledge about one's

likelihood of retrieving material later and about one's previous memory performance is crucial for the optimal allocation of study time and the adjustment of study strategies. Monitoring processes occur during different stages of memory, from stimulus encoding to information retrieval. Specifically, at the time of acquisition, individuals can make judgments of learning (JOLs) to predict the likelihood that they can later retrieve the learned information (Schwartz, 1994; Koriat, 1997). At the time of retrieval, people can be asked to make feeling-of-knowing judgments (FOKs), typically but not always solicited after a failure to recall, to anticipate the chance of remembering the information later (Nelson and Narens, 1990; Koriat, 1993; Schwartz, 1994). After performing a memory task, individuals can make retrospective confidence judgments (RCJs) about the accuracy of their answers (Thompson and Mason, 1996; Brewer and Sampaio, 2012).

Conventionally, these metamemory judgments are probed using rating scales and tasks that alternate between the study and the test (Nelson and Narens, 1990). Extensive research has suggested that how people rate their confidence judgments generally correlate with actual memory performance. This is often defined as a positive Goodman–Kruskal (G–K) gamma correlation between confidence rating and memory accuracy (Nelson and Narens, 1990; Thompson and Mason, 1996; Koriat, 1997). That is, within individuals, items that receive a higher JOL or FOK rating are more likely to be remembered later. Also, answers that receive a higher RCJ rating are more likely to have been incorrect than those with a lower rating. This relative accuracy of judgment, which has been of primary interest in metamemory research, was of focus in the present study.

Different theories have been proposed to explain metamemory judgments and accuracy. Brunswik's (1952) lens model postulates that people use various cues to construct judgments, and hence judgment accuracy is a function of the degree to which the cues being used to construct judgments are predictive of test performance. Similarly, and more recently, Dunlosky and Tauber's isomechanism framework (2013) posits that all metacognitive judgments are based on the same mechanism. What differs is the cues that are available or activated by the type of timing of the judgment being made, which explains why cue utilization and accuracy differ by judgment type. Indeed, JOL (Koriat and Ma'ayan, 2005; Son and Metcalfe, 2005), FOK (Reder, 1987; Koriat, 1993; Metcalfe et al., 1993; Koriat and Levy-Sadot, 2001), and RCJ (Schwartz, 1994; Metcalfe, 2000) have been shown to rely on various mnemonic cues to predict memory performance, but with different weights. In keeping with this view, studies have demonstrated that the accuracies of different types of metamemory judgments are dissociable or minimally interrelated (Kelemen et al., 2000; Souchay and Isingrini, 2012; Le Berre et al., 2016).

Because metamemory skills are important for efficient learning and memory, much research has been devoted to assessing them across populations, including the aged population (Connor et al., 1997; Perrotin et al., 2006). They have also been assessed with respect to neuropsychiatric disorders associated with memory impairments, including Alzheimer's disease

(Souchay, 2007), schizophrenia (Moritz and Woodward, 2006), and autism spectrum disorder (Grainger et al., 2014). Also, neuropsychological studies comparing healthy individuals and lesion patients have suggested a neural basis of metamemory (Pannu and Kaszniak, 2005; Modirrousta and Fellows, 2008; Chua et al., 2014). These studies require that metamemory measures show sufficient consistency (i.e., reproducibility of the position or rank), at least over a short period of time. Indeed, this characteristic determines the value of any psychological test and forms the basis of drawing meaningful conclusions from data about differences between individuals or groups. In addition, these studies would benefit from enhanced knowledge about the stability of metamemory judgment (i.e., preservation of a score) over the course of the task because it elucidates the temporal dynamics and mechanisms of the judgment (e.g., effects of fatigue and learning).

Unfortunately, few studies have evaluated the stability and consistency of the confidence ratings and the accuracy of metamemory judgments, and some have reported low reliability of metamemory accuracy (i.e., gamma scores) over 1–2 weeks. For example, Kelemen et al. (2000) reported significant test–retest correlations for JOL and FOK ratings but negligible test–retest correlations for JOL and FOK accuracies assessed 1 week apart. Similarly, some studies have shown that across two to three sessions that occurred over at least 1–2 weeks, the accuracy of memory and RCJ ratings exhibited moderate to good reliability, whereas RCJ accuracy possessed very low or at most low-to-moderate reliability (Stankov and Crawford, 1996; Thompson and Mason, 1996; Jonsson and Allwood, 2003).

Indeed, although the evidence has been inconclusive and limited in scope, some studies have shown that metamemory accuracy exhibits systematic changes over trials and it lacks consistency, even within a session. Vesonder and Voss (1985) reported a significant improvement in JOL accuracy from the first to the second word list, such that JOL accuracy was above the chance level only for the second list performance. Nelson and Narens (1990) found significantly improved FOK accuracy over three trials of 15 items, but this finding was not replicated in another experiment with four trials of 10 items. Limited by the dearth of evidence, the stability and consistency of RCJ across trials remains unclear. Also, while a distinction has long been made between JOL, FOK, and RCJ (Nelson and Narens, 1990; Kelemen et al., 2000), the stability and consistency of these different types of metamemory judgments have been mainly examined in isolation using different task paradigms. How these features compare among different types of metamemory judgments remains elusive.

The aim of the present study was to evaluate and compare the within-session stability and consistency of JOL, FOK, and RCJ over three trials of paired-associate learning (i.e., episodic metamemory). Each trial consisted of studying a new set of items, followed by a test of memory. While metamemory is commonly studied with word lists or word pairs, face–scene pairs were used in this study to facilitate the use of this metamemory paradigm for future aging

research in Hong Kong. By 2016, 60% of older adults in Hong Kong had attained primary education and below only (Census and Statistics Department, 2018). Thus, using pictorial stimuli, which placed lower demand on literacy than word pairs, would contribute to this research endeavor. In addition, while metamemory accuracy is traditionally indexed by the G–K gamma computed using concordant and discordant pairs of observations, Higham and Higham (2019) recently proposed a new gamma (H–H gamma) that is estimated *via* receiver operating characteristic (ROC) curves and the trapezoidal rule. Simulations have shown that this gamma was more accurate than the traditional gamma in most cases. Thus, both gamma estimates were investigated in this study.

Materials and methods

Participants

A total of 51 Chinese adults aged 18–39 years were originally recruited by means of a poster advertisement on the campus of the Hong Kong Polytechnic University. No participant had a history of any psychiatric or neurological disorder or suffered from a stroke or traumatic brain injury that required hospitalization. None was taking any psychotropic medication. All participants abstained from alcohol and caffeine intake on the day of the experiment. Each one self-reported normal or corrected-to-normal vision. Participants provided written informed consent before being tested individually in a quiet room on the university campus. A total of 13 individuals were excluded from subsequent analysis because they had a missing gamma score in at least one test trial (see Section “Data Analysis”). Therefore, the analytic sample consisted of 38 individuals (17 males, 21 females) with a mean age of 25.5 years ($SD = 5.1$).

The sample size was determined based on previous studies addressing the test–retest or split-half reliability of JOLs, FOKs, and/or RCJs. Across studies, the correlations between confidence ratings at two time points ranged from 0.44 to 0.69 (Thompson and Mason, 1996; Kelemen et al., 2000). Assuming a correlation of 0.44, a power of 0.80, and an alpha level of 0.05, the sample size required to detect a significant correlation was 38. Sample size estimation was not based on the correlations between gamma scores at two time points because they have been found to be very small, if not negligible, which led to impractical sample sizes (Thompson and Mason, 1996; Kelemen et al., 2000). This study was approved by the Human Subjects Ethics Sub-Committee at the Hong Kong Polytechnic University (HSEARS20201110006) and conducted in accordance with the Declaration of Helsinki.

Test materials

A modified face–scene associative learning task was used to probe metamemory (Chua and Solinger, 2015). Test stimuli included 54 photographs of neutral, front-facing Chinese adult

faces taken from the CUHK student database (Wang and Tang, 2008), as well as 54 scenic pictures (18 neutral, 18 positive, and 18 negative) that belonged to the “people” category from the Nencki Affective Picture System (Marchewka et al., 2014). Pictures of different valences were chosen to increase the variability of metamemory ratings and memory performance (Tauber and Dunlosky, 2012). For each individual, the 54 faces and 54 scenes were first randomly paired, and then equally divided into three sets. Each set contained six pictures for each valence. Because there were few pictures for each valence in each trial, and valence had no significant effects on memory performance or metamemory accuracy across items (ANOVA: $p_s > 0.14$), valence was excluded from the subsequent analysis. The IDs of the chosen stimuli are given in the Appendix. The task was programmed using E-Prime 3.0 (Psychology Software Tools, Pittsburgh, PA, United States) and administered on a Lenovo 13.3-inch touchscreen laptop.

Procedure

The metamemory task consisted of three test trials (Figure 1). Each trial began with a study phase, followed by an intervening task and then a test phase (Chua and Solinger, 2015). During the study phase, pairs of faces and scenes were presented on the left and right sides of a computer screen, respectively, one pair at a time. After a face–scene pair was shown for 3 s, the JOL question (“Will you remember after 5 min?”) along with a Likert scale ranging from 1 (“no”) to 9 (“yes”) also appeared on the screen. Participants made a JOL by touching a scale point within 5 s. The face–scene pair remained onscreen for 5 s, for a total of 8 s of study time per pair, followed by an interstimulus interval of 1 s. This design was based on studies that solicited JOL, regardless of a time limit, while studying the stimulus pairs (Kimball and Metcalfe, 2003; Tullis et al., 2013).

After studying all 18 face–scene pairs, participants performed a visual go/no-go distractor task for 60 s. This duration is in keeping with previous studies, in which 30 s (Leonesio and Nelson, 1990; Logan et al., 2012) or 60 s (Modirrousta and Fellows, 2008) was used. During the distractor task, 10 stars, 10 rectangles, 10 triangles, and 10 circles were presented in a trial-by-trial and a randomized order for each individual. Each shape lasted for 0.5 s, followed by an interstimulus interval of 1 s. Participants touched the screen as quickly as possible whenever a star appeared. This task occurred only once in each trial.

The test phase began after the intervening task. For each item, participants were first shown a previously studied face and prompted to indicate whether they could recall the picture paired with the face (i.e., answering “yes” or “no”). Notably, it is impossible to retrieve every detail of a real-world scenic picture studied only for 8 s; indeed, the “yes” and “no” responses reflected the retrievability of partial rather than complete information (see Section “Retrievability of Information and Its Relationship With Metamemory Judgment”). After that, participants were asked to

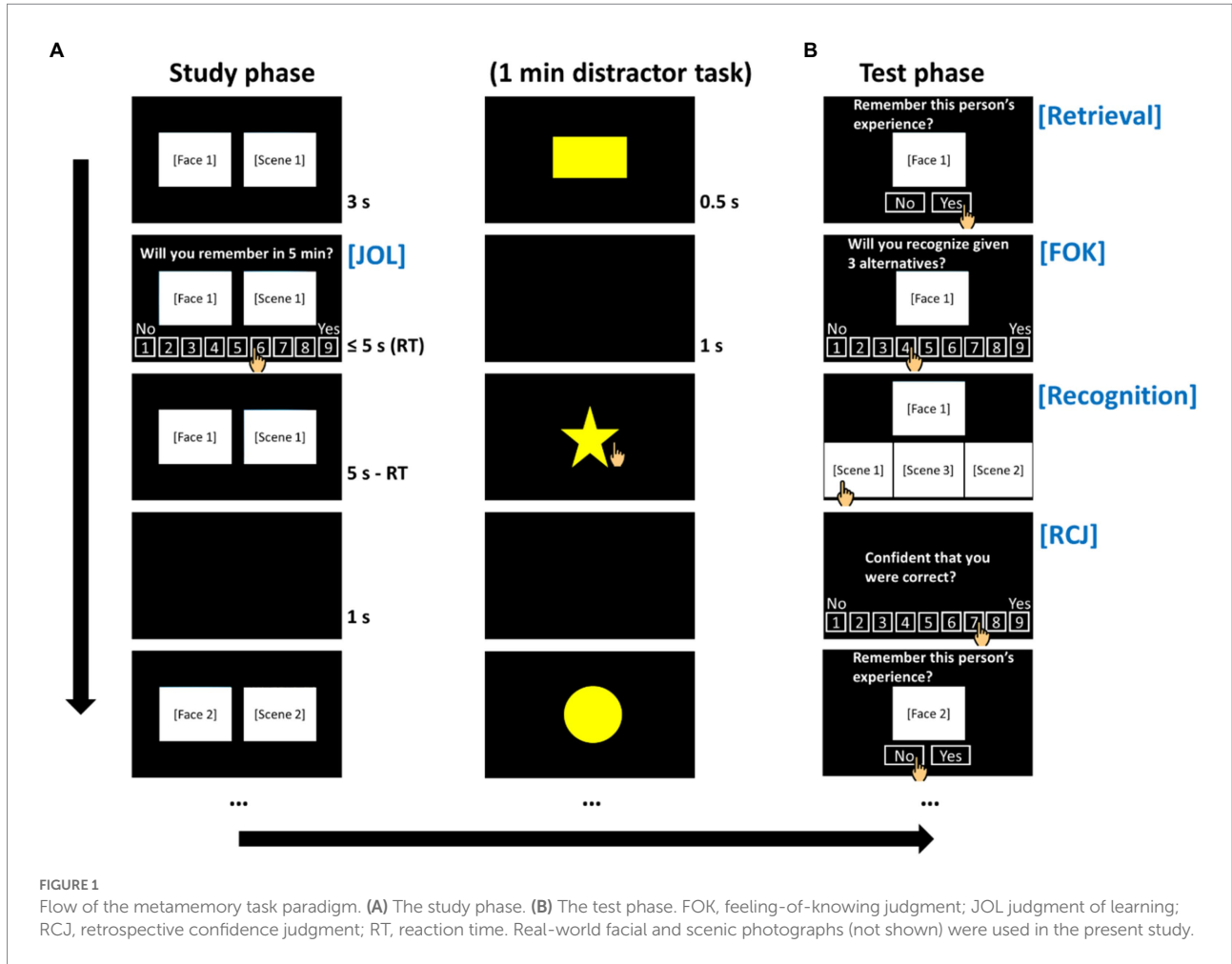


FIGURE 1 Flow of the metamemory task paradigm. (A) The study phase. (B) The test phase. FOK, feeling-of-knowing judgment; JOL judgment of learning; RCJ, retrospective confidence judgment; RT, reaction time. Real-world facial and scenic photographs (not shown) were used in the present study.

make a FOK (“Will you recognize given 3 alternatives?”) by rating on a Likert scale again ranging from 1 (“no”) to 9 (“yes”). In keeping with some previous studies, FOKs were solicited regardless of the outcome of the retrieval attempt (Koriat, 1993; Chua et al., 2009; Watier and Collin, 2011; Chua and Solinger, 2015).

After rating the FOK, participants were given a three-alternative forced-choice recognition test, where a face was presented at the top of the screen and three scenic pictures were shown side by side at the bottom (Modirrousta and Fellows, 2008). One picture was the target, while the other two pictures (i.e., foils) were previously paired with either another same-valence face or a different-valence face. Participants chose the picture that paired with the target face by touching the picture. They then made an RCJ (“Confident that you were correct?”) by rating from 1 (“no”) to 9 (“yes”). The next test trial with a different set of stimuli began after all the face–scene pairs had been rated (using the same scale) and tested.

Before the task started for real, participants were briefed on the task instruction and asked to remember the life experiences of different people presented in the form of photographs (left: face; right: scene). They then underwent a study–test practice trial with

two stimulus pairs (Leonesio and Nelson, 1990). Thus, the participants knew they would receive a recognition test and what they were supposed to predict. Participants were also informed that they would be answering some rating scale questions as they went along. They were instructed to use the full range of the rating scales.

Data analysis

Several analyses were conducted to evaluate the stability and consistency of memory performance, metamemory judgment ratings, and metamemory accuracy across trials. Stability (i.e., change over trials) was evaluated using ANOVAs and *t*-tests, and consistency between and among trials was quantified by calculating Pearson’s correlations and intraclass correlation coefficients (ICCs) (Cicchetti, 1994; Koo and Li, 2016), respectively. Memory performance was evaluated by the proportion of correct answers in the recognition test. Metamemory ratings were based on the mean JOL, FOK, and RCJ ratings. Metamemory accuracy was defined as the gamma correlation between the item-by-item JOL, FOK, or RCJ rating and

recognition performance, which ranges from -1 (perfect negative relationship) to $+1$ (perfect positive relationship).

In the present study, gamma was estimated using two methods. First, the G–K gamma, which has been widely used in metamemory research (Nelson and Narens, 1990), is based on the concept of concordant and discordant pairs (Goodman and Kruskal, 1954).¹ Second, the H–H gamma, which has been recently proposed to be an improved gamma estimate (Higham and Higham, 2019), is computed *via* ROC curves and the trapezoidal rule.² Due to the use of pictorial stimuli and the imperfect relationship between the retrieval response and recognition performance (see Section “Retrievability of Information and Its Relationship With Metamemory Judgment”), FOK accuracy was computed based on all items. While the stability and consistency of the gamma scores were of primary interest, those of the proportion correct and mean ratings also were analyzed to give a context to the gamma scores.

A total of 13 participants had a missing gamma score in at least one trial due to perfect recognition performance or the use of only one rating throughout the entire trial (JOL: $n = 6$; FOK: $n = 9$; RCJ: $n = 8$). Therefore, the analytic sample consisted of 38 participants who provided a complete set of data (i.e., a gamma score for each trial and for each judgment type). Repeated measures ANOVAs were used to assess the stability of the variables. For memory performance, a repeated measures ANOVA with time (trial 1, trial 2, trial 3) as a factor was conducted on the number of correct answers. In addition, two repeated measures ANOVAs with judgment type (JOL, FOK, RCJ) and time as factors were conducted separately on the mean confidence rating and the gamma score. The Greenhouse–Gessier correction was applied when the sphericity assumption was violated.

Pearson’s correlations were calculated to evaluate the consistency of memory performance, metamemory ratings, and metamemory accuracy between trials. Moreover, two-way random effects, consistency ICCs were calculated to evaluate the consistency of the recognition performance, the mean rating, and the gamma scores of each judgment type among all the trials. Consistency is poor for ICCs < 0.40 , fair for ICCs between 0.40 and 0.59 , good for ICCs between 0.60 and 0.74 , and excellent for ICCs ≥ 0.75 (Cicchetti, 1994). Statistical analyses were performed with IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY). All statistical tests were two-tailed, and the alpha level was set at 0.05 . Holm–Bonferroni correction was used to control for inflated familywise error rates because of multiple comparisons. Uncorrected p -values were reported.

1 G–K gamma = $(C - D) / (C + D)$, where C and D refer to the numbers of concordant and discordant pairs, respectively.

2 H–H gamma = $\sum_{k=0}^n (HR_{k+1} + HR_k)(FAR_{k+1} - FAR_k)$, where HR and FAR represent the hit rate and the false alarm rate, respectively; k represents the different criteria plotted on the ROC, and n refers to the number of criteria.

Results

Memory performance

The means and standard deviations of the primary task variables are presented in Table 1, and the Pearson’s correlations and ICCs of the variables are shown in Table 2. First, the stability and consistency of memory performance across the three test trials were examined. A repeated measures ANOVA with time as a factor showed no significant effect of time for the proportion of correct recognition, $p = 0.37$. After Holm–Bonferroni correction, the proportions of correct recognition were significantly positively correlated among all three trials, r_s from 0.38 to 0.56 , $p_s < 0.018$. The single measures ICC was 0.44 , and the average measures ICC was 0.71 , implying good absolute agreement overall.

Metamemory judgment ratings

Next, the stability and consistency of the ratings of metamemory judgments across the trials were analyzed. For stability (Figure 2), a repeated measures ANOVA with judgment type and time as factors was conducted on the mean rating. The main effect of judgment type was significant, $F(1.59, 58.91) = 39.30$, $p < 0.001$, $\eta_p^2 = 0.52$. This was due to a significantly higher rating for RCJ than for JOL and FOK, $p_s < 0.001$. No significant difference in the mean rating between JOL and FOK emerged, $p = 0.33$. Also, the main effect of time was significant, $F(1.61, 59.43) = 3.97$, $p = 0.032$, $\eta_p^2 = 0.097$. This was due to a significant decrease in rating from trial 1 to trial 2, $p = 0.010$, and no significant change thereafter, $p > 0.99$. Importantly, the interaction between judgment type and time was significant, $F(2.81, 103.95) = 8.13$, $p < 0.001$, $\eta_p^2 = 0.18$. Therefore, the effect of time was examined separately for each judgment type. Repeated measures ANOVAs with time as factor revealed a significant result only for JOL, $F(1.53, 56.49) = 14.35$, $p < 0.001$, $\eta_p^2 = 0.28$. The results were not significant for either FOK, $p = 0.67$, or RCJ, $p = 0.42$. For JOL, there was a quadratic decrease in mean rating that diminished over the trials.

Regarding the consistency of metamemory judgment ratings (Figure 3), after Holm–Bonferroni correction, the mean ratings were significantly positively correlated among all three trials for JOL (r_s from 0.64 to 0.88 , $p_s < 0.001$), FOK (r_s from 0.81 to 0.84 , $p_s < 0.001$), and RCJ (r_s from 0.74 to 0.85 , $p_s < 0.001$). Across different types of metamemory judgments, the single measures ICCs ranged from 0.76 to 0.82 , and the average measures ICCs ranged from 0.90 to 0.93 . Therefore, the mean ratings exhibited excellent consistency across types of metamemory judgment.

Metamemory accuracy

The stability and consistency of metamemory accuracy across trials were then analyzed. Regarding stability (Figure 4), a repeated measures ANOVA with judgment type and time as

TABLE 1 Means and standard deviations of the memory performance, metamemory judgment, and metamemory accuracy variables ($n=38$).

Variable	Test trials					
	Trial 1		Trial 2		Trial 3	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Memory performance						
Recognition accuracy (%)	71.1	15.0	73.5	13.9	74.4	14.2
Metamemory judgment						
Mean JOL rating	5.5	1.6	4.8	1.6	4.6	1.6
Mean FOK rating	5.2	1.4	5.1	1.4	5.1	1.5
Mean RCJ rating	6.1	1.6	6.1	1.7	6.3	1.6
Metamemory accuracy						
JOL G–K gamma	–0.11	0.47	0.15	0.46	0.31	0.51
FOK G–K gamma	0.28	0.47	0.38	0.42	0.30	0.51
RCJ G–K gamma	0.43	0.50	0.55	0.42	0.60	0.27
JOL H–H gamma	–0.22	0.43	0.06	0.39	0.19	0.40
FOK H–H gamma	0.14	0.38	0.26	0.30	0.21	0.37
RCJ H–H gamma	0.22	0.50	0.39	0.46	0.34	0.44

FOK, feeling of knowing; G–K, Goodman–Kruskal; H–H, Higham–Higham; JOL, judgment of learning; RCJ, retrospective confidence judgment.

TABLE 2 Pearson's correlations and intraclass correlation coefficients (ICCS) of the memory performance, metamemory judgment, and metamemory accuracy measures among the three test trials ($n=38$).

Variable	Pearson's correlation (<i>r</i>)			ICC	
	Trial 1–Trial 2	Trial 2–Trial 3	Trial 1–Trial 3	Single measures	Average measures
Memory performance					
Recognition accuracy (%)	0.56***	0.39*	0.38*	0.44 [0.25, 0.63]	0.71 [0.50, 0.84]
Metamemory judgment					
Mean JOL rating	0.74***	0.88***	0.64***	0.76 [0.63, 0.85]	0.90 [0.83, 0.95]
Mean FOK rating	0.83***	0.84***	0.81***	0.82 [0.72, 0.90]	0.93 [0.89, 0.96]
Mean RCJ rating	0.78***	0.85***	0.74***	0.79 [0.68, 0.88]	0.92 [0.86, 0.96]
Metamemory accuracy					
JOL G–K gamma	0.10	0.34*	0.00	0.15 [–0.04, 0.37]	0.34 [–0.13, 0.64]
FOK G–K gamma	0.07	0.18	0.16	0.14 [–0.05, 0.36]	0.33 [–0.15, 0.63]
RCJ G–K gamma	0.06	0.47**	0.23	0.20 [0.00, 0.42]	0.42 [0.01, 0.68]
JOL H–H gamma	0.23	0.42**	0.28	0.30 [0.10, 0.52]	0.57 [0.26, 0.76]
FOK H–H gamma	–0.14	0.19	0.23	0.10 [–0.08, 0.32]	0.25 [–0.28, 0.59]
RCJ H–H gamma	0.57***	0.60***	0.36*	0.50 [0.31, 0.68]	0.75 [0.58, 0.86]

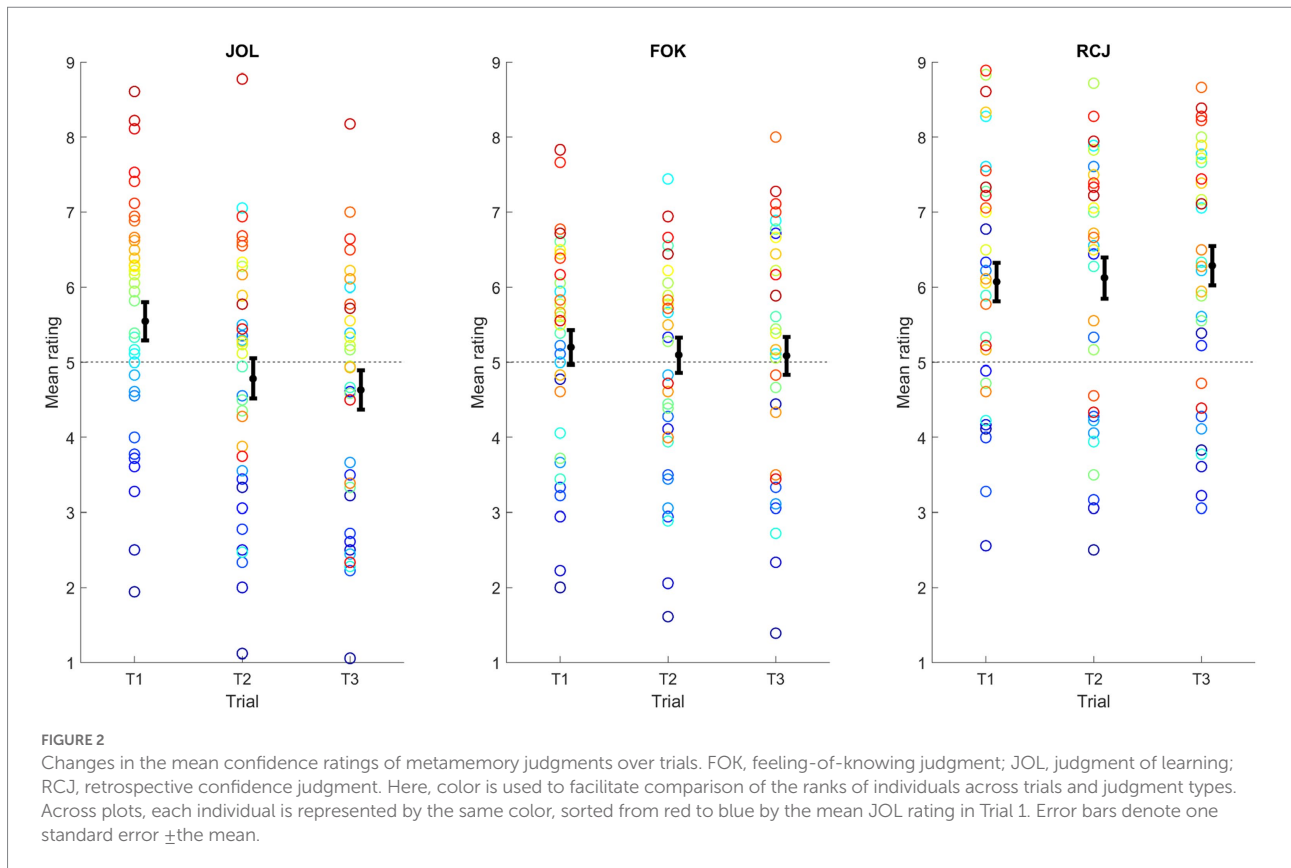
FOK, feeling-of-knowing judgment; G–K, Goodman–Kruskal; H–H, Higham–Higham; JOL, judgment of learning; RCJ, retrospective confidence judgment. The ICCs were based on two-way random effects models with consistency. Bold values survived Holm–Bonferroni correction.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

factors conducted on the G–K gamma score showed a significant main effect of judgment type, $F(2, 74) = 19.14$, $p < 0.001$, $\eta_p^2 = 0.34$. This was due to a significantly larger gamma score for RCJ than for JOL and FOK, and for FOK than for JOL, $ps < 0.003$. There was also a significant main effect of time, $F(2, 74) = 5.42$, $p = 0.006$, $\eta_p^2 = 0.13$. This was owing to a significant linear increase in the gamma score over trials, $p = 0.004$. Importantly, the interaction between judgment type and time also was significant, $F(4, 148) = 2.77$, $p = 0.030$, $\eta_p^2 = 0.070$. Hence, the effect of time was analyzed by conducting a repeated measures ANOVA with time as a factor for each judgment type.

The repeated measures ANOVAs identified a significant main effect of time only for JOL, $F(2, 74) = 8.49$, $p < 0.001$, $\eta_p^2 = 0.19$. It was not significant for either FOK, $p = 0.54$, or RCJ, $p = 0.14$. This result was attributable to a significant linear increase in the JOL gamma score over trials, $p < 0.001$. Indeed, after Holm–Bonferroni correction, one-sample t -tests showed that all but two gamma scores were significantly larger than zero and that the JOL gamma score did not significantly predict recognition performance until the third trial (Trial 1 JOL: $p = 0.16$; Trial 2 JOL: $p = 0.054$; Others: $ps < 0.001$).

Albeit with lower magnitudes overall, the H–H gamma scores yielded almost the same ANOVA and t -test results. That is,



ANOVA still revealed significant main effects of judgment type, $F(1.6, 59.7) = 10.79$, $p < 0.001$, $\eta_p^2 = 0.23$, and time, $F(2, 74) = 8.85$, $p < 0.001$, $\eta_p^2 = 0.19$. There also was a significant interaction between judgment type and time, $F(4, 148) = 3.69$, $p = 0.007$, $\eta_p^2 = 0.09$, which was driven by a significant linear increase over trials for JOL, $p < 0.001$, but not for others, $ps > 0.17$. Most H–H gamma scores, particularly those representing the last trial, $ps < 0.006$, were significantly above zero.

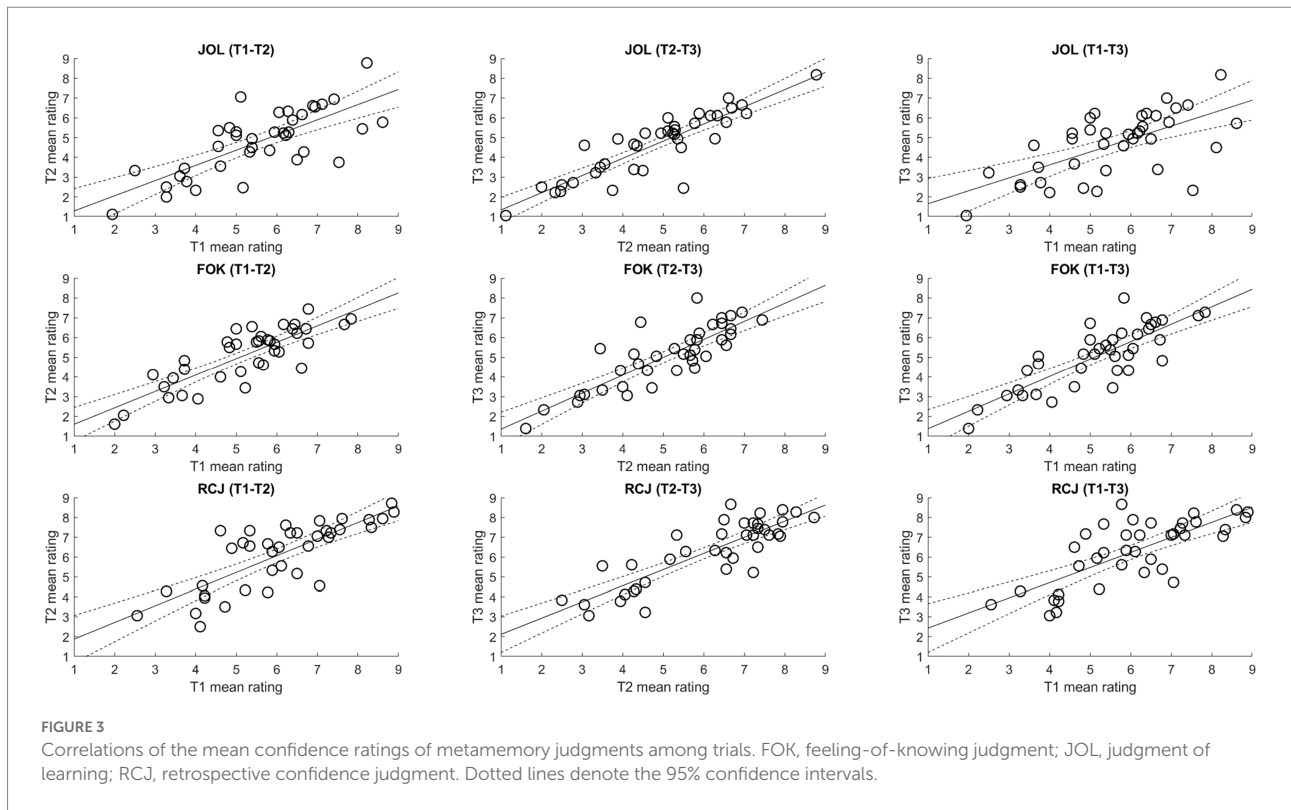
Regarding the consistency of gamma scores, the G–K gamma scores generally showed poor consistency over the trials (Figure 5). Specifically, after Holm–Bonferroni correction, there was only a marginally significant positive correlation between trial 2 and trial 3 JOL gamma scores, $r(36) = 0.34$, $p = 0.034$, and a significant positive correlation between trial 2 and trial 3 RCJ gamma scores, $r(36) = 0.47$, $p = 0.003$ (others: ps from -0.003 to 0.23 , $ps > 0.17$). Across types of metamemory judgment, the average measures ICCs ranged from 0.30 to 0.42, implying poor consistency overall.

In contrast, the consistency of H–H gamma scores varied greatly across different types of metamemory judgment (Figure 6). The correlation between trial 2 and trial 3 JOL gamma scores and the correlations among the three trials' RCJ gamma scores were (marginally) significant, $rs > 0.36$, $ps < 0.028$. In addition, the single measures ranged from 0.10 to 0.50, and the average measures ICCs for JOL, FOK, and RCJ were 0.57, 0.25, and 0.75, suggesting fair-to-good, poor, and good-to-excellent consistency, respectively.

Retrievability of information and its relationship with metamemory judgment

Additional analyses were performed to clarify the mechanisms underlying the metamemory processes, and the results are shown in Table 3. Before making a FOK, the participants were prompted to indicate whether they could retrieve the scenic picture associated with the face (i.e., making a retrieval attempt). To clarify the nature of the retrieved information, the mean FOK ratings and the proportions of correct recognition following a “yes” and a “no” response were compared by conducting two separate repeated measures ANOVAs with retrievability (yes, no) and time as factors. Unfortunately, only 24 (10 males, 14 females) out of the 38 participants allowed for this analysis, as they gave a mix of “yes” and “no” responses on each of the three trials, with no significant difference in the proportion of items that received a “yes” answer among trials, $p = 0.46$. Nevertheless, comparing those who were subsequently included and excluded did not reveal a significant difference in any variable, including the proportion of items with a “yes” response, $p > 0.16$. Also, none of the previously reported results significantly changed after repeating the analyses (e.g., ANOVAs) with this subset of the sample. Thus, these 24 individuals appear to be representative of the entire study sample.

For the mean FOK rating, a significant result was obtained for the main effect of retrievability, $F(1, 23) = 108.64$, $p < 0.001$, $\eta_p^2 = 0.83$. However, the main effect of time was not significant,



$p=0.44$. The interaction between retrievability and time was marginally significant, $F(2, 46) = 3.18$, $p = 0.051$, $\eta_p^2 = 0.12$. *Post-hoc t*-tests revealed a significant difference between the “yes” and “no” responses on each trial, $ps < 0.001$. The interaction was due to a significant linear increase in the difference in mean FOK rating between the two responses over trials, $p = 0.030$. In addition, for the proportion of correct recognition, the main effect of retrievability was significant, $F(1, 23) = 22.33$, $p < 0.001$, $\eta_p^2 = 0.49$. This owed to a more accurate recognition performance following the “yes” than the “no” response. No other effect was significant, $ps > 0.20$. Notably, the mean FOK ratings and recognition performances following the “yes” and “no” responses were far from extreme, suggesting that the two responses likely reflected the retrieval of different amounts of partial information rather than of the complete presence and absence of information.

To elucidate whether the rating of metamemory judgment correlated with the amount of retrieved information cued by a face, a repeated measures ANOVA with judgment type and time as factors was conducted on the G–K gamma correlation between the item-by-item rating and retrievability. The main effect of time was significant, $F(2, 46) = 4.82$, $p = 0.013$, $\eta_p^2 = 0.17$. This could be understood as a significant linear increase in the gamma score over trials, $p = 0.014$. The main effect of the judgment type also was significant, $F(2, 46) = 53.67$, $p < 0.001$, $\eta_p^2 = 0.70$. This was driven by a larger FOK gamma score than the other two gamma scores, and by a larger RCJ gamma score than the JOL gamma score, $ps < 0.001$. No significant interaction between judgment type and time emerged, $p = 0.52$. Using the H–H gamma gave similar results

(Time: $p = 0.018$; Judgment Type: $p < 0.001$; Time \times Judgment Type: $p = 0.34$).

Discussion

Metamemory is important for selecting and applying learning strategies that promote memory (Nelson and Narens, 1990). Therefore, there has been much interest in studying and assessing metamemory skills across healthy and clinical populations (Connor et al., 1997; Perrotin et al., 2006; Souchay, 2007; Grainger et al., 2014). The assessment of metamemory relies on the premise that measures of metamemory skills are consistent across repeated measurements; otherwise, these measures cannot be used to infer the true ability of a person. Also, the stability or change of metamemory judgment over the course of the task is important to consider since it informs the temporal dynamics and mechanisms of the judgment. Unfortunately, these aspects have rarely been systematically or comprehensively evaluated. The aim of the present study was to bridge this knowledge gap by comparing the within-session stability and consistency of three major types of metamemory judgments using a single-task paradigm.

The present study focused on gamma (relative accuracy) that was estimated using two methods: the traditional method that considers concordant and discordant pairs (i.e., G–K gamma; Goodman and Kruskal, 1954), and a recent method that is based on area under the ROC curve (i.e., H–H gamma; Higham and Higham, 2019). Regardless of the measure, there was a significant

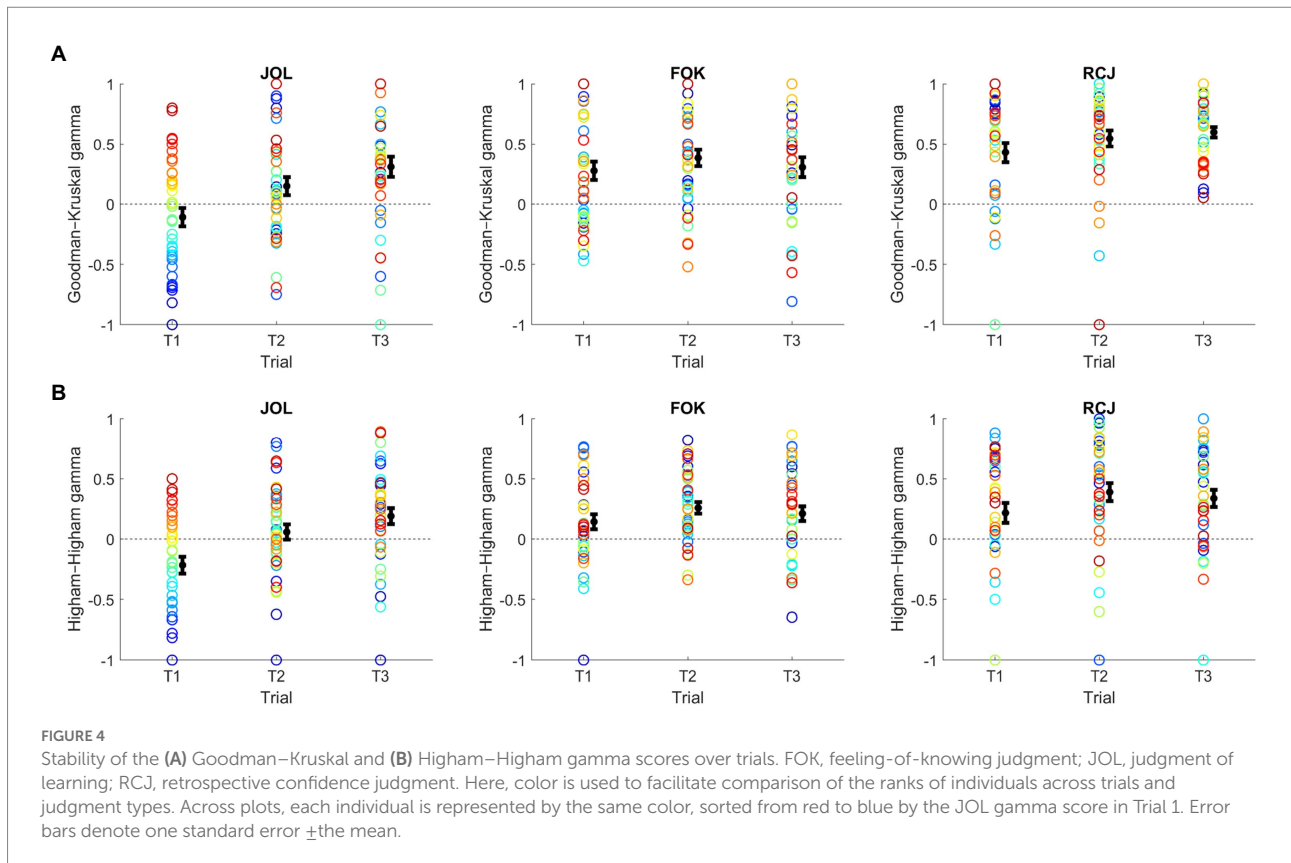


FIGURE 4

Stability of the (A) Goodman–Kruskal and (B) Higham–Higham gamma scores over trials. FOK, feeling-of-knowing judgment; JOL, judgment of learning; RCJ, retrospective confidence judgment. Here, color is used to facilitate comparison of the ranks of individuals across trials and judgment types. Across plots, each individual is represented by the same color, sorted from red to blue by the JOL gamma score in Trial 1. Error bars denote one standard error \pm the mean.

improvement in JOL accuracy but not in FOK or RCJ accuracy over the trials. This selective change was qualified by a significant interaction between judgment type and time. This finding corroborates a previous finding of a significant increase in JOL accuracy from the first to the second trial of list learning, such that JOL ratings did not significantly predict memory performance until the latter stage of the task (Vesonder and Voss, 1985).

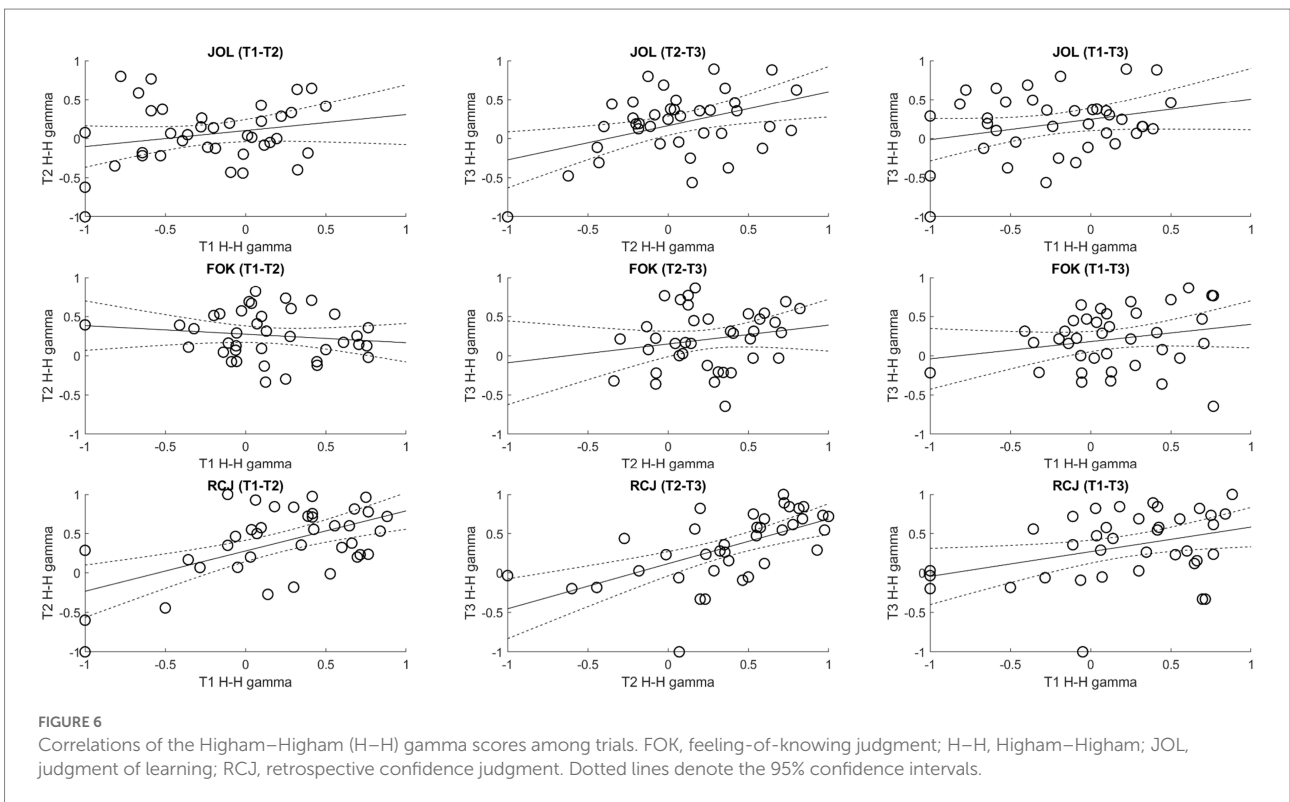
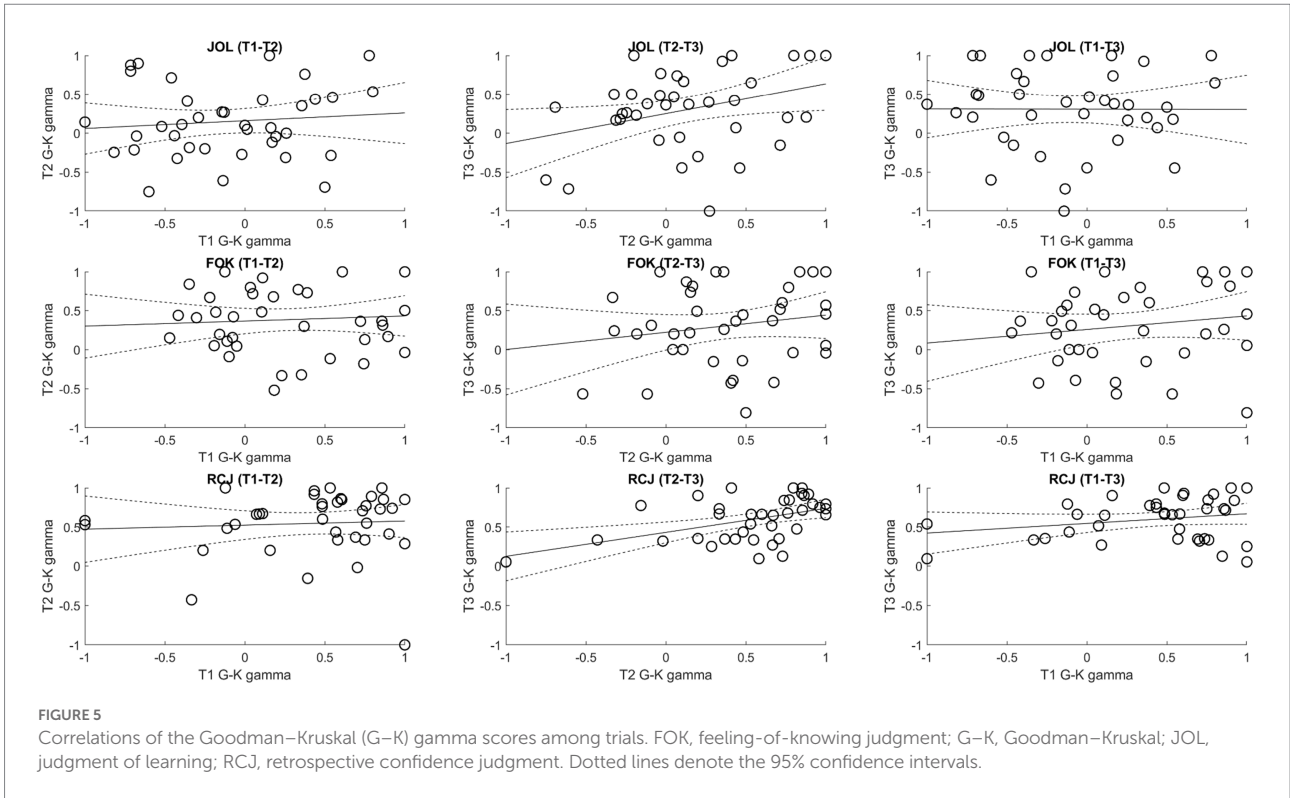
In addition, the current finding is consistent with the inconclusive change in FOK accuracy within sessions. Specifically, Nelson and Narens (1990) found that FOK accuracy gradually improved over three test trials in one experiment, but this finding was not replicated in another experiment. To the best of my knowledge, no study has examined the stability of RCJ accuracy within sessions. Nevertheless, like the present study, one study reported no significant change in RCJ accuracy over 2 weeks (Jonsson and Allwood, 2003). Taken together, the present study yields novel evidence that among the three major types of metamemory judgment, only JOL accuracy exhibits systematic changes over successive trials.

According to the isomechanism theory (Dunlosky and Tauber, 2013), all metacognitive judgments are based on the same mechanism. However, since the availability and activation of cues differ by the type of timing of the judgment being made, cue utilization and accuracy differ by judgment type. In the present study, the decrease in JOL ratings over trials suggests that the participants might be adjusting their expectations of their own memory performance based on their performance on prior

memory tests (Finn and Metcalfe, 2007). In addition, the improved JOL accuracy over time may imply that the participants learned something about the task and applied it to changing how they made their JOLs going forward. Because a significant improvement in accuracy was observed only for JOL, the adjustment and learning appear to be specific to cue utilization at the time of studying. For example, valence could be an obvious cue during JOL (Tauber and Dunlosky, 2012), but it did not predict memory performance in the present study. The participants might change from relying on valence to other (diagnostic) cues when making JOLs over the course of the task.

Recognition performance and mean confidence ratings were found to exhibit good to excellent consistency across types of metamemory judgment. The H–H gamma scores for JOL and RCJ also demonstrated at least fair-to-good consistency. Thus, although the gamma scores were defined as the item-by-item relationships between confidence ratings and memory performance, the poor consistency of some gamma scores, particularly FOK gamma scores, cannot be simply attributable to unreliable confidence ratings or memory performance. In addition, the relatively small number of items (i.e., 18) per trial, although still larger than that in some studies (Nelson and Narens, 1990; Pinon et al., 2005; Logan et al., 2012), is unlikely to explain the inconsistency in these scores.

Indeed, the consistency of judgment accuracy was found to vary by the type of metamemory judgment and by the gamma estimation method. For JOL and RCJ, the H–H gamma showed



improved consistency compared to the G–K gamma. For FOK, however, neither gamma measure yielded consistent results. Higham and Higham (2019) performed a series of simulations

and found that the ROC-based gamma deviated less from the true value of gamma than did the traditional gamma in most of the simulations (Higham and Higham, 2019). Therefore, the higher

TABLE 3 Means and standard deviations of variables related to the retrieval response, nature of the retrieved information, and the relationship between the ratings and retrieval responses ($n=24$).

Variable	Test trials					
	Trial 1		Trial 2		Trial 3	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Retrieval response						
Proportion of answering “yes” (%)	47.1	33.2	47.1	33.9	44.3	33.0
Nature of the retrieved information						
Mean FOK rating after answering “yes”	6.8	1.3	6.7	1.3	7.0	1.0
Mean FOK rating after answering “no”	4.3	1.3	4.1	1.3	3.8	1.4
Recognition accuracy after answering “yes” (%)	81.5	18.3	86.3	17.8	83.6	17.9
Recognition accuracy after answering “no” (%)	70.0	18.2	73.4	15.1	65.6	20.2
Relationship between the ratings and retrieval responses						
JOL G–K gamma	0.09	0.51	0.29	0.36	0.37	0.44
FOK G–K gamma	0.77	0.44	0.80	0.27	0.96	0.09
RCJ G–K gamma	0.58	0.35	0.63	0.29	0.69	0.32
JOL H–H gamma	0.00	0.46	0.21	0.34	0.31	0.41
FOK H–H gamma	0.73	0.30	0.73	0.29	0.88	0.15
RCJ H–H gamma	0.32	0.47	0.41	0.40	0.39	0.40

FOK, feeling of knowing; G–K, Goodman–Kruskal; H–H, Higham–Higham; JOL, judgment of learning; RCJ, retrospective confidence judgment.

accuracy of the ROC-based gamma may contribute to the enhanced consistency of this new measure, at least for JOL and RCJ. In addition, based on various theories of metacognitive judgment (Brunswik, 1952; Dunlosky and Tauber, 2013), the observed wide range of consistencies across the new gamma scores may indicate that cue utilization undergoes varying extents of idiosyncratic trial-by-trial fluctuation depending on the stage of judgment about learning and memory.

Metamemory judgments are well known to be based on a combination of intrinsic, extrinsic, and mnemonic cues (Koriat, 1997; Koriat and Levy-Sadot, 2001; Koriat and Ma'ayan, 2005; Son and Metcalfe, 2005). Koriat (1997) showed that JOL accuracy improved with practice, and the improved resolution was associated with an increased reliance on mnemonic–experiential cues rather than on intrinsic cues. The present task involved the random pairing of faces and scenes that were different for each individual. Therefore, a change in the reliance on intrinsic cues, such as item relatedness, is unlikely to underlie the selective improvement in JOL accuracy over trials.

The item-by-item correlation between confidence ratings and the retrievability of information significantly increased over trials. This implies an increased reliance on the accessibility of information at the time of judgment. However, this increased correlation was not significantly moderated by the type of judgment. Therefore, how the increase in using this mnemonic cue is linked with the specific improvement in JOL accuracy over trials remains elusive. In addition, as in most metamemory studies, the time interval between metamemory judgment and memory performance was longest for JOL. Therefore, an enhanced use of extrinsic cues (e.g., task structure) that became increasingly apparent over practice, as well as a more accurate estimation of the

time interval between the study and the test, may have the greatest impact on JOL accuracy.

In some studies, the G–K gamma scores exhibited poor test–retest reliability over 1–2 weeks. Kelemen et al. (2000) reported significant correlations for JOL and FOK ratings but not for the corresponding gamma scores between two sessions that were 1 week apart. Also, Thompson and Mason (1996) found that the levels of split-half and alternate-forms reliability were moderate-to-good for memory performance and median confidence ratings. However, they were poor for RCJ accuracy assessed on two occasions that took place over a two-week interval (also see Stankov and Crawford, 1996; Jonsson and Allwood, 2003). The present study addressed the within-session stability and consistency of metamemory judgments, and the findings suggest that the poor reliability of these gamma scores can also be detected within one test session. Considering the improved within-session consistency of the ROC-based gamma, future work would benefit from evaluating the reliability of this new measure as applied to evaluating the accuracy of metamemory judgments across sessions.

The present study has several notable strengths. First, it employed a within-subjects design and compared three major types of metamemory judgments using a single, ecologically valid task paradigm. This approach allowed for the direct comparison of different kinds of metamemory judgments with minimal confounds. Second, this study was one of the first to systematically evaluate the within-session stability and reliability of different metamemory variables. It contributes to the literature by revealing the temporal dynamics of metamemory judgments and by clarifying the psychometric properties of several commonly used metamemory metrics. Third, it pioneered to evaluate and compare the stability and consistency of two different gamma estimates. In light of the

current findings, the ROC-based gamma is recommended over the traditional gamma for evaluating the relative accuracy of metamemory judgments, particularly JOL and RCJ.

Notwithstanding these contributions, this study has some limitations. First, it addressed only the monitoring aspect of metamemory. The stability and consistency of the control processes of metamemory, including the allocation of study time and the selection of search strategies, remain unclear. Second, ease-of-learning judgments, which have been considered a core monitoring process of metamemory that takes place prior to acquisition (Nelson and Narens, 1990), were not examined. Third, this study did not provide information about the stability and consistency of metamemory judgment ratings and accuracies over repeated exposure to the same set of study items (Koriat, 1997). Fourth, the present results were based on healthy young adults, and they might not be generalizable to other age groups or clinical populations.

In summary, this study examined and compared the within-session stability and consistency of three major types of metamemory judgments using a single-task paradigm. The results showed that only JOL accuracy improved over trials, and this observation was paralleled by a selective decrease in mean JOL ratings. Depending on the type of metamemory judgment and the choice of gamma estimate, the within-session consistency of metamemory accuracy varied greatly, ranging from poor to excellent. This study provides preliminary support for using the ROC-based gamma over the traditional gamma to evaluate the relative accuracy of judgment. It also highlights the need for more research on the temporal dynamics of metamemory judgment and on the psychometric properties of metamemory accuracy measures to improve the understanding and assessment of metamemory abilities.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by The Human Subjects Ethics Sub-Committee at the Hong Kong Polytechnic University (HSEARS20201110006). The

References

Brewer, W. F., and Sampaio, C. (2012). The metamemory approach to confidence: a test using semantic memory. *J. Mem. Lang.* 67, 59–77. doi: 10.1016/j.jml.2012.04.002

Brunswik, E. (1952). *The Conceptual Framework of Psychology*. Vol. 1. Chicago, IL: University of Chicago Press.

Census and Statistics Department (2018). Thematic report: older persons. Available at: https://www.byccensus2016.gov.hk/data/16BC_Older_persons_report.pdf (Accessed August 8, 2022).

patients/participants provided their written informed consent to participate in this study.

Author contributions

MY contributed to study conception and design, data analysis and interpretation, and manuscript writing.

Funding

This study was supported by the Start-up Fund for RAPs under the Strategic Hiring Scheme (P0034754) and by the One-Line Budget (P0039302) awarded by the Hong Kong Polytechnic University to MY.

Acknowledgments

The author would like to thank Vivian Chu for data collection.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.917576/full#supplementary-material>

Chua, E. F., Pergolizzi, D., and Weintraub, R. R. (2014). "The cognitive neuroscience of metamemory monitoring: understanding metamemory processes, subjective levels expressed, and metacognitive accuracy," in *The Cognitive Neuroscience of Metacognition*. eds. S. Fleming and C. Frith (Berlin, Heidelberg: Springer), 267–291.

Chua, E. F., Schacter, D. L., and Sperling, R. A. (2009). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. *J. Cogn. Neurosci.* 21, 1751–1765. doi: 10.1162/jocn.2009.21123

- Chua, E. F., and Solinger, L. A. (2015). Building metamemorial knowledge over time: insights from eye tracking about the bases of feeling-of-knowing and confidence judgments. *Front. Psychol.* 6:1206. doi: 10.3389/fpsyg.2015.01206
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol. Assess.* 6, 284–290. doi: 10.1037/1040-3590.6.4.284
- Connor, L. T., Dunlosky, J., and Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychol. Aging* 12, 50–71. doi: 10.1037/0882-7974.12.1.50
- Dunlosky, J., and Tauber, S. K. (2013). “Understanding people’s metacognitive judgments: an isomechanism framework and its implications for applied and theoretical research,” in *The Sage Handbook of Applied Memory*. eds. T. J. Perfect and D. S. Lindsay. (Thousand Oaks, CA: SAGE Publications Ltd), 444–464.
- Finn, B., and Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *J. Exp. Psychol. Learn. Mem. Cogn.* 33, 238–244. doi: 10.1037/0278-7393.33.1.238
- Goodman, L. A., and Kruskal, W. H. (1954). Measures of association for cross classifications. *J. Am. Stat. Assoc.* 49, 732–769. doi: 10.2307/2281536
- Grainger, C., Williams, D. M., and Lind, S. E. (2014). Metacognition, metamemory, and mindreading in high-functioning adults with autism spectrum disorder. *J. Abnorm. Psychol.* 123, 650–659. doi: 10.1037/a0036531
- Higham, P. A., and Higham, D. P. (2019). New improved gamma: enhancing the accuracy of Goodman–Kruskal’s gamma using ROC curves. *Behav. Res. Methods* 51, 108–125. doi: 10.3758/s13428-018-1125-5
- Jonsson, A. C., and Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personal. Individ. Differ.* 34, 559–574. doi: 10.1016/S0191-8869(02)00028-4
- Kelemen, W. L., Frost, P. J., and Weaver, C. A. (2000). Individual differences in metacognition: evidence against a general metacognitive ability. *Mem. Cogn.* 28, 92–107. doi: 10.3758/BF03211579
- Kimball, D. R., and Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Mem. Cogn.* 31, 918–929. doi: 10.3758/BF03196445
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychol. Rev.* 100, 609–639. doi: 10.1037/0033-295X.100.4.609
- Koriat, A. (1997). Monitoring one’s own knowledge during study: a cue-utilization approach to judgments of learning. *J. Exp. Psychol. Gen.* 126, 349–370. doi: 10.1037/0096-3445.126.4.349
- Koriat, A., and Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychol. Rev.* 103, 490–517. doi: 10.1037/0033-295X.103.3.490
- Koriat, A., and Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 34–53. doi: 10.1037/0278-7393.27.1.34
- Koriat, A., and Ma’ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *J. Mem. Lang.* 52, 478–492. doi: 10.1016/j.jml.2005.01.001
- Le Berre, A. P., Müller-Oehring, E. M., Kwon, D., Serventi, M. R., Pfefferbaum, A., and Sullivan, E. V. (2016). Differential compromise of prospective and retrospective metamemory monitoring and their dissociable structural brain correlates. *Cortex* 81, 192–202. doi: 10.1016/j.cortex.2016.05.002
- Leonesio, R. J., and Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 464–470. doi: 10.1037/0278-7393.16.3.464
- Logan, J. M., Castel, A. D., Haber, S., and Viehman, E. J. (2012). Metacognition and the spacing effect: the role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacogn. Learn.* 7, 175–195. doi: 10.1007/s11409-012-9090-3
- Marchewka, A., Żurawski, Ł., Jednoróg, K., and Grabowska, A. (2014). The Nencki affective picture system (NAPS): introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behav. Res. Methods* 46, 596–610. doi: 10.3758/s13428-013-0379-1
- Metcalfe, J. (2000). “Metamemory: theory and data,” in *The Oxford Handbook of Memory*. eds. E. Tulving and F. I. M. Craik (London: Oxford University Press), 197–211.
- Metcalfe, J., Schwartz, B. L., and Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 851–861. doi: 10.1037/0278-7393.19.4.851
- Mudirrousta, M., and Fellows, L. K. (2008). Dorsal medial prefrontal cortex plays a necessary role in rapid error prediction in humans. *J. Neurosci.* 28, 14000–14005. doi: 10.1523/JNEUROSCI.4450-08.2008
- Moritz, S., and Woodward, T. S. (2006). The contribution of metamemory deficits to schizophrenia. *J. Abnorm. Psychol.* 115, 15–25. doi: 10.1037/0021-843X.115.1.15
- Nelson, T. O., and Narens, L. (1990). Metamemory: a theoretical framework and new findings. *Psychol. Learn. Motiv.* 26, 125–173. doi: 10.1016/S0079-7421(08)60053-5
- Pannu, J. K., and Kaszniak, A. W. (2005). Metamemory experiments in neurological populations: a review. *Neuropsychol. Rev.* 15, 105–130. doi: 10.1007/s11065-005-7091-6
- Perrotin, A., Isingrini, M., Souchay, C., Clarys, D., and Taconnat, L. (2006). Episodic feeling-of-knowing accuracy and cued recall in the elderly: evidence for double dissociation involving executive functioning and processing speed. *Acta Psychol.* 122, 58–73. doi: 10.1016/j.actpsy.2005.10.003
- Pinon, K., Allain, P., Kefi, M. Z., Dubas, F., and Le Gall, D. (2005). Monitoring processes and metamemory experience in patients with dysexecutive syndrome. *Brain and Cognition* 57, 185–188. doi: 10.1016/j.bandc.2004.08.042
- Reder, L. M. (1987). Strategy selection in question answering. *Cogn. Psychol.* 19, 90–138. doi: 10.1016/0010-0285(87)90005-3
- Schwartz, B. L. (1994). Sources of information in metamemory: judgments of learning and feelings of knowing. *Psychon. Bull. Rev.* 1, 357–375. doi: 10.3758/BF03213977
- Son, L. K., and Metcalfe, J. (2005). Judgments of learning: evidence for a two-stage process. *Mem. Cogn.* 33, 1116–1129. doi: 10.3758/BF03193217
- Souchay, C. (2007). Metamemory in Alzheimer’s disease. *Cortex* 43, 987–1003. doi: 10.1016/S0010-9452(08)70696-8
- Souchay, C., and Isingrini, M. (2012). Are feeling-of-knowing and judgment-of-learning different? Evidence from older adults. *Acta Psychologica* 139, 458–464. doi: 10.1016/j.actpsy.2012.01.007
- Stankov, L., and Crawford, J. D. (1996). Confidence judgments in studies of individual differences. *Personal. Individ. Differ.* 21, 971–986. doi: 10.1016/S0191-8869(96)00130-4
- Tauber, S. K., and Dunlosky, J. (2012). Can older adults accurately judge their learning of emotional information? *Psychol. Aging* 27, 924–933. doi: 10.1037/a0028447
- Thompson, W. B., and Mason, S. E. (1996). Instability of individual differences in the association between confidence judgments and memory performance. *Mem. Cogn.* 24, 226–234. doi: 10.3758/BF03200883
- Tullis, J. G., Finley, J. R., and Benjamin, A. S. (2013). Metacognition of the testing effect: guiding learners to predict the benefits of retrieval. *Mem. Cogn.* 41, 429–442. doi: 10.3758/s13421-012-0274-5
- Vesonder, G. T., and Voss, J. F. (1985). On the ability to predict one’s own responses while learning. *J. Mem. Lang.* 24, 363–376. doi: 10.1016/0749-596X(85)90034-8
- Wang, X., and Tang, X. (2008). Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 31, 1955–1967. doi: 10.1109/TPAMI.2008.222
- Watier, N. N., and Collin, C. A. (2011). Metamemory for faces, names, and common nouns. *Acta Psychol.* 138, 143–154. doi: 10.1016/j.actpsy.2011.05.017