What’s the Best Way to Measure Global Metacognition?

A Comparison Between Task-Congruent and Incongruent Beliefs

Rohini Sen
University of Chicago

A paper submitted in partial fulfillment of the requirements for the
Master of Arts degree in the
Master of Arts Program in the Social Sciences
June 2024

Faculty Advisor: Dr. David Gallo, Ph.D.

Preceptor: Dr. Hannah Hamilton, Ph.D.

Acknowledgements

This thesis would not have reached its full potential without the expertise and guidance of my faculty advisor, Dr. David Gallo. His deep and extensive knowledge of memory and cognitive neuroscience of aging has been invaluable. I am immensely grateful to Dr. Gallo for the confidence and direction he provided, both of which were crucial for undertaking this project. I must also express my deep appreciation for Dr. Hannah Hamilton, my program preceptor, whose unwavering support and invaluable resources have profoundly shaped the completion of this thesis. Her thoughtful guidance has not only facilitated my work but also enriched my entire research experience.

I owe a profound debt of gratitude to my graduate student advisor, Tesnim Arar, whose unwavering support and insightful feedback were essential throughout the research and analysis stages of this thesis. From the conceptual framework to the final draft, her encouragement to explore new ideas and her meticulous edits were invaluable. Tesnim's guidance was instrumental in shaping this thesis into its successful final form.

I am grateful to the University of Chicago for recognizing my abilities and providing an ideal academic environment that fostered my growth through nuanced and comprehensive learning experiences.

I express heartfelt gratitude to my parents, who believed in my potential even before I could see it myself and have supported me with patience and love. I am grateful to my sister for being my first role model and mentor, continually inspiring me with her prowess in learning and her love for writing. Lastly, I thank my friends worldwide, whose care and companionship have not only empowered and supported me throughout my academic journey but also filled it with inspiration and joy.

Abstract

Despite being the focus of countless investigations in recent decades, metacognition as a construct remains poorly understood. This is especially true for one subset of metacognition—global metacognition—which comprises generalized assessments of one's cognitive abilities. One major hindrance to the advancement of theoretical frameworks on global metacognition is the absence of appropriate measurement techniques. Using a novel cognitive battery and global metacognition questionnaire (Metacognitive Insight Questionnaire or MIQ; Wong & Gallo, 2019), recent work from our lab (Arar et al., in prep) has attempted to address this methodological limitation by introducing an objective way to measure global metacognition. In addition to providing this new technique, this work also sought to examine how task experience may impact people's global metacognitive beliefs. However, this work did not consider the potential conflation of item types on the MIQ, specifically items that are more pertinent to our cognitive battery, or *congruent*, with items that are *incongruent*. The present work addresses this limitation; here, we sought to investigate whether task-incongruent and congruent beliefs are differentially affected by task exposure in younger and older adults. Results revealed that younger adults exhibited task-induced belief updating *only* for task-congruent beliefs, demonstrating that younger adults may be able to adaptively utilize relevant cues to inform and update their global metacognitive beliefs. Older adults, however, did not exhibit this discrimination effect; they adjusted both belief types downward (i.e., expressed less confidence in their cognitive abilities) after completing our cognitive battery, perhaps illustrating stereotype threat effects. Altogether, these findings offer important insights into the best ways to measure global metacognition and further our understanding of the updating and formation of global metacognitive beliefs.

Measuring Global Metacognitive Accuracy:

A Comparison Between Task-Congruent and Task-Incongruent Beliefs

Metacognition, or knowledge of one's cognitive abilities and processes (also known informally as "thinking about thinking"), is central to adaptive behavior. Possessing a high degree of metacognitive accuracy—or congruency between one's subjective metacognitive beliefs and objective cognitive ability—has been linked to optimal self-regulation and superior performance on a variety of tasks (e.g., maximizing goal-related memory outcomes and performance; McGillivray & Castel, 2017). Given this relationship, some researchers have postulated that enhancing the accuracy of one's metacognitive beliefs is an antecedent to improving cognitive performance and behavior (Hertzog & Dunlosky, 2011; Castel, 2018). However, not all metacognitive beliefs are equal in scope and magnitude. Indeed, metacognitive beliefs span many levels of abstraction, ranging from item-level or task-specific judgments, or local metacognitive beliefs, to more overarching assessments, or global metacognitive beliefs.

Of specific interest to us are global metacognitive beliefs, which are thought to be most influential for everyday behavior (Rouault & Fleming, 2020), a notion supported by work demonstrating the impact of global metacognitive beliefs on financial decision making (Yu et al., 2022) and risky behavior (Starkstein et al., 2007). Indeed, such work suggests that global metacognition may be an ideal target for interventions aimed at improving everyday behaviors, which may be especially useful for older adults experiencing age-related cognitive decline (see Hertzog & Dunlosky, 2011; Hargis & Castel, 2018).

It is unfortunate then that global metacognition remains critically understudied, at least, outside of a clinical context. To date, most of the literature aimed at measuring global metacognition has been in the context of clinical assessment of cognitive decline, as in the case of Mild Cognitive Impairment (MCI) or Alzheimer's disease (AD). Here, a typical procedure to measure global metacognitive insight or accuracy is to administer questionnaires to participants to have them rate their own cognitive abilities (or severity of cognitive problems), and then to compare these ratings to those of an informant (usually a significant other or clinician) in order to identify the extent of the discrepancy (e.g., Clare, 2004; Leicht, Berwig, & Gertz, 2010; Starkstein et al., 2006). Although this technique has produced several important findings, a major limitation is that this technique lacks an objectively anchored assessment of cognitive ability from which to assess global metacognitive accuracy.

Beyond the difficult task of creating a cognitive test that could both gauge global cognitive ability and serve to compute an objective measure of global metacognitive ability, researchers would also need to ensure that any measure is uncontaminated by confounding influences such as their general tendency to give lower or higher confidence ratings, also known as metacognitive bias (McWilliams, 2023). This endeavor poses several challenges, primarily because factors coming to bear on global metacognition are largely unknown and, perhaps more importantly, any measure of global metacognition would need to adequately isolate metacognitive ability from intertwined cognitive processes (Peters, 2022). Often, studies measuring global metacognitive abilities may conflate measures or ideas of metacognitive bias, metacognitive sensitivity and metacognitive efficiency (Fleming, 2014).

Given the complexity of validating measures of global metacognition or studying it in the context of healthy cognitive aging, two initial questions must first be addressed. The first: What is the appropriate way to measure global metacognition? A second: What are the effects of normative aging on global metacognition? Despite the value of intact metacognitive processes in helping older adults compensate for age-related declines in cognitive performance, few researchers have attempted to address this question. Recent findings by McWilliams et al. (2023) on the effects of age on metacognitive judgements in cognitively healthy adults suggest that our capacity to accurately assess task performance, known as local metacognitive efficiency, remains stable across our lifespan. This observation aligns with previous studies that have consistently reported intact local metacognitive abilities in cognitively healthy older adults (Bertrand et al., 2017; Halamish et al., 2011; Hertzog et al., 2002; Robinson et al., 2006; Rouault, Seow, et al., 2018; Souchay et al., 2007). However, whether global metacognition is similarly spared from age-related cognitive decline is unknown. Although, McWilliams et al. (2023) did obtain what they refer to as a "global metric of metacognition", it should be noted that these judgments were specific to the cognitive task employed in their study, and thus, perhaps not a true measure of global metacognition (at least, in the traditional sense).

The present work seeks to address these two questions, building on prior work from our lab (Arar et al., in prep), by exploring the impact of aging and cognitive task experience on global and local metacognitive beliefs. Initial findings from Arar et al. suggests that both cognitively normal younger and older adults demonstrated initial overconfidence in their cognitive abilities before taking our novel cognitive battery (from here on, referred to as the "CAT-COG"). Upon completion of the CAT-COG, older adults reevaluated their global metacognitive beliefs, such that their initial overconfidence was reduced, thereby leading to more calibrated global metacognitive beliefs. A notable limitation of this research, however, is that due to the conflation of MIQ items that are more directly pertinent to the cognitive battery, or what we refer to as *task congruent items,* with items that are not, or *task incongruent* Inventory Questionnaire (MIQ), it is difficult to discern whether recent cognitive task exposure affects all global metacognitive judgments equally, or, whether task-induced belief updating occurs only for task-congruent MIQ items. For example, evaluating global metacognitive beliefs about semantic memory involved participants rating their ability to identify common objects (a process aligned with the object naming task in the cognitive battery) alongside their proficiency in understanding the meanings of uncommon words (a broader everyday life skill that isn't directly aligned with the cognitive task on the battery).

Thus, the present research aims to address this limitation and disentangle the effects of cognitive task exposure on these different belief types by conducting a re-analysis of data from Arar et al. (in prep). Although all items on the MIQ reflect broader everyday cognitive behaviors, this reanalysis aims to distinguish between global metacognitive judgments closely linked to specific cognitive battery tasks and those reflecting behaviors within that specific cognitive domain but not directly aligned with the cognitive processes assessed in our tasks.

Method

Participants

Participants (*N* = 637) were recruited from the online platform Prolific (https://www.prolific.com/) and compensated for their participation in the experiment. Participants were binned into one of six age groups (18-29, 30-49, 50-64, 65-75, 76-86-year-olds). All participants were pre-screened to ensure they were between 18-86 years old, fluent English speakers, had normal to corrected-to-normal vision, were not color-blind, did not have untreated mental disorders, and had not been diagnosed with Alzheimer's disease (AD) or Mild Cognitive Impairment (MCI; an intermediate state between normal cognitive aging and dementia). For this thesis, only data from two of the age groups were examined/ Thus, our final sample included 196 younger adults (115 female; age range: 18-29 years; M_{age} = 23.76; SD_{age} =

3.17) and 196 older adults (57 female; age range = 65-75 years; M_{age} = 68.22; SD_{age} = 2.96). There were no significant differences between the younger and older adults on account of any other demographic characteristics, including mean years of formal education, all *t*s < 1.13, all *p*s > 0.26. The demographic details of these participants are depicted in **Table 1**.

Condition	**Pre-Post**		**Only Post**	
Age Group	Younger Adults	Older Adults	Younger Adults	Older Adults
Total *N* (*n* females)	98 (56)	97 (56)	98 (59)	99 (53)
Age				
Mean (*SD*)	23.74 (3.21)	68.42 (3.05)	23.78 (3.14)	68.02 (2.87)
Race/Ethnicity				
White/Caucasian	61.2%	96.9%	57.1%	99.0%
Black	13.3%	—	19.4%	1.0%
Asian	17.4%	2.06%	17.4%	—
Other	5.1%	1.03%	4.10%	—
Multiracial	2.0%	—	2.0%	—
Unknown or not reporting	—	—	—	—
Education				
Mean years of education (*SD*)	15.6 (2.25)	15.5 (3.0)	14.1 (5.1)	14.8 (3.5)
High school degree or equivalent	38.8%	26.8%	30.6%	28.3%
Associate's/Technical degree	4.1%	16.5%	7.14%	12.12%
Undergraduate degree	46.0%	32.0%	46.0%	33.3%
Master's degree	10.2%	15.5%	13.3%	14.14%
Doctorate (Ph.D.)	—	2.06%	1.02%	4.04%
Professional degree (e.g., M.D.)	—	4.12%	—	—
No formal diploma	1.0%	3.10%	2.04%	8.1%
Handedness				
Right	80.6%	84.5%	87.76%	91.0%
Left	17.4%	14.4%	12.25%	6.1%
Ambidextrous	2.0%	1.03%	—	3.03%

Table 1. Demographic information and number of participants per condition (not including participants that were excluded from data analysis). Standard deviations are included in parentheses.

Procedure

The study was conducted online and consisted of a single session for both younger and older adults. Participants were recruited via Prolific, an online recruitment platform, and were placed into one of two conditions that differed solely on how many times participants took the MIQ. In one condition, which we refer to as the “Pre + Post Condition”, participants took the MIQ both before and immediately after the CAT-COG. Whereas in the “Only Post Condition,” participants took the MIQ only once, at the end of the session. Upon enrolling in the study, participants in both conditions were immediately redirected to Qualtrics, an online survey and data collection platform, where they provided informed consent, took the MIQ (if applicable), and provided demographic information. Next, all participants took the cognitive battery and then received the MIQ questionnaire (i.e., the post-task MIQ). The order of experimental tasks is shown below in **Figure 1**. Please note that for this study, only data from younger adults (18-29-year-olds) and older adults (65-75-year-olds) in the “Pre + Post condition” (*N* = 195) have been analyzed.

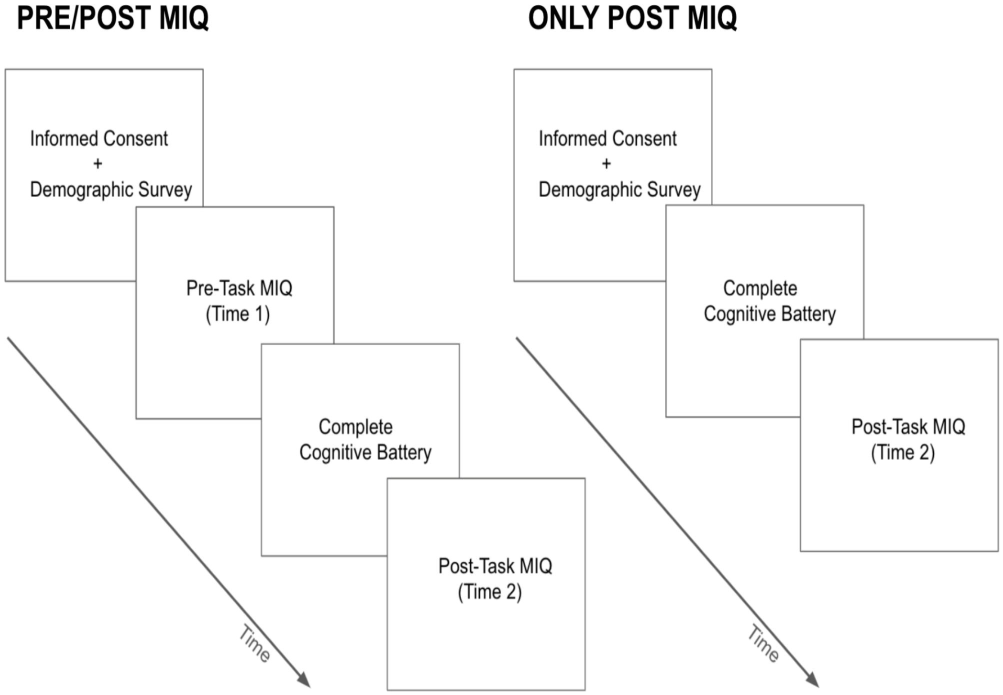

Figure 1. Experimental Task Order. In Condition 1, participants receive both pre-task and post-task MIQ (Pre/Post). In Condition 2, participants receive only the post-task MIQ (Only Post). MIQ = Memory Insight Questionnaire.

Measures

Cognitive Battery (i.e., "CAT-COG")

To assess global cognitive ability, we used our novel computerized cognitive test battery, or the "CAT-COG", designed to assess the five following domains of cognitive functioning: episodic memory, semantic memory, working memory, executive function, and processing speed. These domains are typically assessed in neuropsychological tests and are considered most relevant to complex mental functioning. The battery consisted of three test blocks, organized by increasing difficulty (i.e., easy block, medium block, hard block). The tasks assessing each of the following cognitive subdomains are described below. Cognitive performance in each domain (e.g., episodic memory) was calculated as the proportion of correct responses on the tasks described below. The scores from all five domains were then averaged to derive a global cognitive ability measure. These raw scores were converted into Z scores to assess cognitive performance within each domain.

Episodic Memory (Recognition Memory Task). For assessing episodic memory, a recognition memory task was conducted with either an immediate recognition task or a delayed recognition task at the end of the complete battery. Participants were shown nonsense words, one at a time for 1000 ms, with an interstimulus interval (ISI) of 1000 ms. These words were strings of letters from a nonsense word generator. Following the encoding phase, participants were asked to solve distractor True/False arithmetic problems during a 12 s retention interval, evaluating whether equations such as "3 + 8 = 11" were mathematically correct. During the test, participants were presented with a set of four options (four pictures or four words, based on the modality of the target item) and were instructed to identify which option they had previously encountered. The difficulty of the trials varied; for instance, more challenging picture recognition trials featured images of the same object type with subtle differences (e.g., a cream-colored bike from the encoding phase might be among choice options of cream-colored bikes with slightly different features), whereas difficult nonsense word trials included options phonetically close to the target, and challenging real word trials offered semantically related options. The delayed recognition memory task consisted of nine trials conducted at the end of the cognitive battery, wherein participants were shown nine-word targets from the earlier short-delay memory recognition task and were asked to recognize the stimuli.

Semantic Memory (Object Naming Task). We used an object identification task to assess participants' semantic memory ability. During this task, participants were presented with images of animate (e.g., fish) and inanimate (e.g., vase) objects similar to those used in Konkle

& Caramazza (2013). The task contained 27 trials of varying difficulty, spread equally across the battery's three test blocks. It featured 27 target pictures, distinct from those in the episodic recognition memory task, depicting real-world objects and not illustrations. Participants had four response options to choose from, with the response time being self-paced. In more challenging trials, the options provided were semantically related to the target object (e.g.,), whereas easier trials offered semantically distant choices (e.g.).

Working Memory (Forward and Backward Digit Span Task). For assessing working memory, participants engaged in both forward and backward digit span tasks, with sequences ranging from three to ten digits for forward tasks and three to eight for backward tasks. The forward digit span contained 15 trials, whereas the backward contained 12. During both tasks, individual numbers were displayed for 1000 ms, followed by a 1000 ms interstimulus interval. After the last digit in the sequence was presented, there was either a 1000 ms, 3000 ms, or 5000 ms delay (depending on difficulty level), after which participants were asked to provide the digit sequence in either the same (forward) order or in reverse (backward) order, again with the same delay options and response time.

Executive Functioning (Color Matching Task). We used two tasks to evaluate executive functioning. The first was a modified version of the Stroop task, where participants were presented with a list of color words written in different ink colors (e.g., blue, purple, green) and were asked to count how many of the words were displayed in colors that matched their semantic meaning. The second was an evaluation task we refer to as the "rule identification task". Here, participants viewed sets of objects that varied in shape and color and had a number from 1-5 in their center. Two to five objects were shown at a time, and participants were asked to identify the common feature among the set (color, shape, or number). Both tasks, designed to measure different aspects of executive function, consisted of 27 trials evenly distributed across the test blocks.

Processing Speed (Fast Comparison Task). To assess processing speed, participants engaged in a string comparison task, where they were presented with pairs of strings containing numbers, letters, and symbols (e.g., BKK and KK3, including symbols like $). They had to evaluate whether the strings were identical or different, with strings varying in length from two to seven characters based on trial difficulty. Participants were instructed to respond as quickly as possible, with each trial having a timeout of 5000 ms.

Memory Insight Questionnaire (MIQ)

To assess participants' global metacognitive beliefs, we used a novel global metacognitive measure that we refer to as the Metacognitive Insight Questionnaire (MIQ; Wong & Gallo, 2019). The MIQ featured 10 items, with two MIQ items for each cognitive subdomain evaluated in the cognitive battery. These items covered the same five cognitive domains—episodic memory, semantic memory, working memory, executive functioning and processing speed—assessed by our cognitive battery. Participants rated their abilities compared to the average person their age across various abilities, such as recalling new names, places, or grocery lists, on a 7-point scale (1 = *substantially worse*; 7 = *substantially better*). For an overall measure of global metacognitive insight, the responses to all 10 questionnaire items were averaged. Additionally, responses on items for each subdomain were averaged separately, providing five distinct measures of metacognitive insight per participant (e.g., episodic memory insight). Although all items probe general beliefs about one’s cognition, half are particularly relevant to the cognitive battery used in this study, potentially making those five items more susceptible to task-induced belief updating. Take, for example, the two items pertaining to working memory.

The first probed participants' metacognitive beliefs about their ability to rehearse phone numbers and directions, and the second participants' perceived ability to multitask (e.g., cooking while talking). In this case, the first item is most related to the working memory task in our cognitive battery (i.e., digit span). As such, the current study differentiates between items that are task congruent and incongruent and aims to investigate whether these global metacognitive beliefs are differentially affected by task exposure. The MIQ items, the domains they relate to, and the differentiation between items congruent with cognitive tasks on the battery and those not congruent are outlined in **Table 2**.

MIQ Questionnaire Item	**Domain**	**Cognitive Battery Task**
Recalling new names, places, or grocery lists.	Episodic Memory	Congruent Task: Memory Recognition
Remembering that you've already told someone a story.		Incongruent
Identifying common objects.	Semantic Memory	Congruent Task: Object Identification and Naming
Knowing the meaning of uncommon words.		Incongruent
Rehearsing phone numbers or directions.	Working Memory	Congruent: Forward and Backward Digit Span Task
Multi-tasking (e.g., cooking while talking).		Incongruent
Paying close attention (e.g., reading in a noisy room).	Executive Functioning	Congruent: Color Matching Task (similar to the Stroop Task)
Learning new tasks (e.g., new card games).		Congruent: Feature Matching Task (Rule Identification)
Reacting quickly (e.g., hitting the brakes when driving).	Processing Speed	Congruent: Fast Comparison Task
Performing speeded tasks.		Incongruent

Table 2. MIQ questionnaire items, categorized by their task congruence (i.e., whether they are CAT-COG-congruent or incongruent beliefs) based on their congruence with specific tasks in the cognitive battery. The MIQ contains 10 items total, with two items per the five cognitive domains examined. Note: Executive functioning was evaluated through two distinct tasks, one focused on color matching and the other on rule identification. For these tasks, an MIQ item considered congruent with one task was considered incongruent with the other and vice versa.

Metacognitive Accuracy

To evaluate participants' global metacognitive accuracy—how closely their subjective beliefs about cognitive abilities align with objective performance—the following methods were used in this study. First, participants' proportion correct on each of the different CAT-COG subtasks were transformed into Z scores. These scores served as an objective measure of cognitive ability. Finally, to compute a measure of global metacognitive accuracy, participants' CAT-COG Z scores were subtracted from the MIQ scores. A schematic representation of this accuracy calculation is shown in **Figure 2** below. Participants' metacognitive ratings per domain can be found in **Table 2** and participants' metacognitive accuracy scores per domain can be found in **Table 3** (see **Results** section).

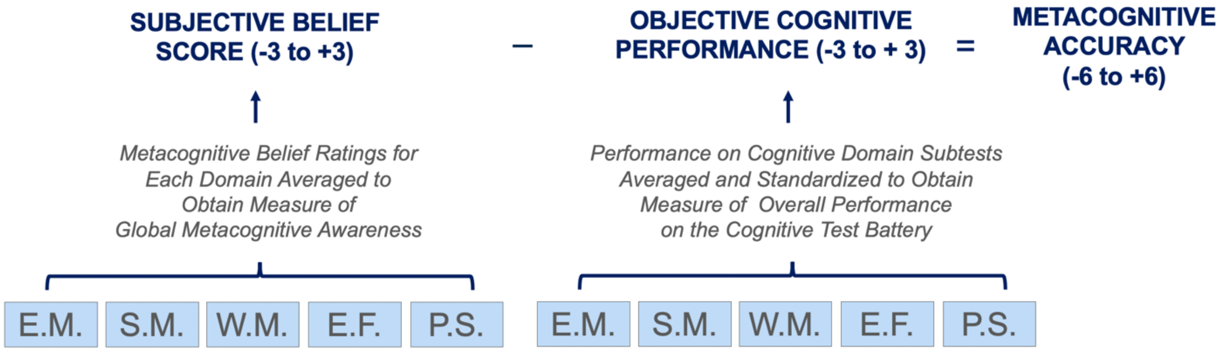

Figure 2. Calculation of metacognitive accuracy variable. Objective global cognitive performance (transformed into a z-score; range = -3 to +3) was subtracted from global metacognitive awareness (average rating on Memory Insight Questionnaire; range = -3 to +3) to compute metacognitive accuracy (range = -6 to +6). E.M = Episodic Memory, S.M. = Semantic Memory, W.M. = Working Memory, E.F. = Executive Function, P.S. = Processing Speed.

Results

Effects of Task Congruence & Time Within Age Groups on Global Metacognitive Ratings

Younger Adults. 2 (congruence: task-congruent vs. incongruent) x 2 (time: pre- and post-task) ANOVA was conducted to examine differences in younger adults' task-congruent and task-incongruent MIQ items at two time points: pre- and post-cognitive battery assessment. The ANOVA indicated a significant main effect of congruence, $F(1, 97) = 14.64$, $p < .001$, $\eta^2_p = .13$.

However, a main effect of time was not found (although, some may argue that the effect is marginally significant), *F*(1, 97) = 3.11, *p* = .081, η^2_p = .03, replicating findings from Arar et al. (in prep). Importantly, however, there was a significant interaction between congruence and time, *F* (1, 97) = 5.01, *p* = .03, η^2_p = .05, indicating that task experience does differentially affect younger adults' global metacognitive beliefs depending on how congruent those beliefs are with the task. To decompose this interaction, follow-up paired *t* tests were conducted and revealed that younger adults rated themselves higher on task-congruent MIQ items (*M* = 0.49; *SD* = 0.85) after taking the cognitive battery (i.e., at time 2) compared to before taking the battery (*M* = 0.37; *SD* = 0.84), *t*(97) = -2.54, *p* = .013, *d* = 0.15. The same was not true for incongruent-MIQ items; younger adults' ratings did not significantly differ from pre-task (*M* = 0.57; *SD* = 0.81) to post-task (*M* = 0.60; *SD* = 0.88), *t*(97) = -0.62, *p* = .54, Cohen's *d* = -0.06. Younger adults' pre-assessment and post-assessment task-congruent and incongruent ratings can be found in **Table 4** and **Table 5** respectively. For comparison's sake, pre- and post-cognitive task MIQ ratings collapsed across congruency (i.e., Arar et al.'s initial findings) for each age group can be found in **Table 3**.

Older Adults. A 2 (congruence: task-congruent vs. incongruent) x 2 (time: pre- and post-task) ANOVA was conducted to examine differences in older adults' task-congruent and task-incongruent MIQ items at two time points: pre- and post-cognitive battery assessment. The ANOVA indicated a significant main effect of congruence, *F* (1, 97) = 97.23, *p* < .001, η^2_p = .50. A significant main effect of time was also found, *F* (1, 97) = 4.52, *p* = .04, η^2_p = .05, bolstering findings from Arar et al. (in prep). However, the interaction between congruence and time was not significant, *F*(1, 97) < .01, *p* = 0.98, η^2_p < 0.001, suggesting that task experience does not differentially affect older adults' global metacognitive beliefs based on the congruence of those beliefs with the task. To decompose this interaction, follow-up paired *t* tests were conducted and revealed that older adults rated themselves lower on task-congruent MIQ items after taking the cognitive battery (i.e., at time 2) (*M* = 0.17; *SD* = 0.66) compared to before taking the battery (*M* = 0.24; *SD* = 0.58), but this difference was not statistically significant, *t(*96) = -1.97, *p* = 0.051, *d* = 0.16. For incongruent-MIQ items, older adults' ratings did not significantly differ from pre-task (*M* = 0.58; *SD* = 0.61) to post-task (*M* = 0.51; *SD* = 0.64), *t*(96) = 1.83, *p* = .07, *d* = 0.15. Older adults' pre-assessment and post-assessment task-congruent and incongruent ratings can be found in **Table 6** and **Table 7** respectively.

	Age Group			
	Younger Adults		**Older Adults**	
Time Point	**Pre-Test**	**Post-Test**	**Pre-Test**	**Post-Test**
Total *N* (*n* females)	98 (56)	98 (56)	97 (56)	97 (56)
MIQ Ratings (*SD*)				
Global Score	0.49 (0.77)	0.55 (0.84)	0.43 (0.55)	0.35 (0.61)
Episodic Memory	0.58 (1.02)	0.57 (0.88)	0.21 (0.71)	0.25 (0.73)
Semantic Memory	0.73 (0.89)	0.71 (0.99)	0.96 (0.72)	0.84 (0.74)
Working Memory	0.28 (1.15)	0.41 (1.07)	0.38 (0.77)	0.16 (0.86)
Executive Function	0.37 (1.02)	0.54 (0.98)	0.32 (0.80)	0.30 (0.84)
Processing Speed	0.52 (1.05)	0.50 (1.18)	0.26 (0.74)	0.20 (0.79)

Table 3. Pre- and post-cognitive battery MIQ ratings–collapsed across task congruency–for younger and older adults. Standard deviations included in parentheses. Global score reflects the average of MIQ ratings across domains.

Effects of Congruence and Age on Global Metacognitive Ratings

Time 1. To gauge the relationship between age and congruence more directly, baseline (i.e., time 1) MIQ ratings were submitted to a 2 (age group: younger vs. older) by 2 (congruence: task-congruent vs. incongruent) mixed ANOVA. A significant main effect of congruence was found, $F(1, 193) = 73.66$, $p < .001$, $\eta^2_p = 0.28$, with participants rating themselves significantly lower on task-congruent beliefs ($M = 0.31$; $SD = 0.73$) compared to task-incongruent beliefs ($M = 0.58$; $SD = 0.72$), $t(194) = -8.49$, $d = -0.60$. This finding shows that task congruence significantly influenced MIQ ratings across age groups at baseline. The interaction between congruence and age group, $F(1, 193) = 4.576$, $p = .03$, $\eta^2p = .02$, did not meet a stricter significance threshold of $p = .01$. No significant main effect of age alone was observed at baseline, $F(1, 193) = 0.42$, $p = .52$, $\eta^2p = .002$.

Time 2. To compare age and congruence post-assessment (i.e., time 2) MIQ ratings were submitted to a 2 (age group: younger vs. older) by 2 (congruence: task-congruent vs. incongruent) mixed ANOVA. At post-assessment, the mixed ANOVA continued to show a significant main effect of congruence, $F(1, 193) = 65.16$, $p < .001$, $\eta^2p = .25$, with task-congruent beliefs rated lower ($M = 0.34$; $SD = 0.77$) than task-incongruent beliefs ($M = 0.56$; $SD = 0.77$), $t(194) = -7.74$, $p < .001$, when aggregating the MIQ ratings from both younger and older adults. There was no main effect of age, $F(1, 193) = 3.68$, $p = .06$, $\eta^2_p = .02$. Here again, a significant interaction between congruence and age group was found, $F(1, 193) = 16.981$, $p < .001$, $\eta^2_p = .08$. To decompose this interaction, paired t tests were conducted to compare the differences between older adults and younger adults on the task-congruent and incongruent MIQ items, respectively. For the congruent items, the difference between the MIQ ratings of the 18-29-year-olds ($M = 0.49$, $SD = 0.85$) and the 65-75-year-olds ($M = 0.18$, $SD = 0.66$) was statistically significant ($p = .004$). However, for the incongruent items, there was no statistical difference between the ratings from the 18-29-year-olds (M =0.60, SD =0.88) and the 65-75-year-olds (M =0.51, SD =0.64).

Effects of Congruence and Domain on Global Metacognitive Ratings

Younger Adults. To examine the relationship between cognitive domain and congruence, two 6 (domain: episodic memory, semantic memory, working memory, executive function–Stroop, executive function–rule identification, processing speed) x 2 (congruence: task-congruent vs. incongruent) repeated measures ANOVA was applied to MIQ ratings (one ANOVA to examine baseline ratings and the other to examine post-task ratings).

With respect to the pre-task/baseline MIQ results, there was a significant main effect of congruence, $F(1, 97) = 16.1$, $p < .001$, with a large effect size ($\eta^2_p = 0.14$). Domain exhibited a significant main effect, $F(5, 485) = 5.71$, $p < .001$, with a small to moderate effect size ($\eta^2_p = .06$), with the highest global metacognitive ratings observed in the domain of semantic memory ($M = 0.73$; $SD = 0.89$), and the lowest ratings observed in the domain of working memory ($M = 0.28$; $SD = 1.15$). Additionally, there was a significant interaction between congruence and domain, $F(5, 485) = 10.8$, $p < .001$, with a large effect size ($\eta^2_p = 0.10$). However, the violation of the sphericity assumption for this interaction may affect the interpretation of these results. To decompose this interaction, follow-up paired t tests comparing task-congruent and task-incongruent ratings across the six domains were conducted. Some significant differences did emerge when breaking down the MIQ ratings by cognitive domain. All ratings can be found in **Table 4**, but most notably, when younger adults evaluated their episodic and working memory abilities, we do see a difference between task-congruent versus task-incongruent items. They rated themselves significantly lower on congruent ($M = 0.33$; $SD = 1.11$) compared to incongruent (*M* = 0.81; *SD* = 1.15) episodic memory items, *t*(97)= -3.35, *p* = .001, *d* = -0.34; as well as significantly lower on congruent (M = -0.16; SD = 1.46) compared to incongruent (M = 0.72; SD = 1.38) working memory items, *t*(97) = -5.29, p < .001, *d* = -0.53. The same was not found for the domains of executive function, semantic memory, and processing speed. A Bonferroni correction with a cutoff score of α = 0.0042 was applied to account for the increased risk of Type I error due to multiple comparisons.

For the post-test analysis, a significant main effect was observed for congruence, *F*(1, 97) = 7.37, *p* = .008, η^2_p = .07), but a non-significant effect for domain, *F*(5, 485) = 2.08, *p* = .07, η^2_p = .02). Furthermore, the significant interaction effect between congruence and domain, *F*(5, 485) = 8.59, *p* < .001, η^2_p = .08) persisted. Examining younger adults' post-task MIQ ratings, significant differences between congruent versus incongruent items emerged only in working memory, with younger adults rating themselves significantly lower on congruent (*M* = 0.06, *SD* = 1.38) compared to incongruent (*M* = 0.77, *SD* = 1.31) items. A Bonferroni correction with a cutoff score of α = 0.0042 was applied. Younger adults' post-task MIQ ratings, broken down by congruence can be found in **Table 5.**

Older Adults. With respect to the pre-task/baseline MIQ results, there was a significant main effect of congruence*,* *F*(1, 97) = 82.7, *p* < .001, with a large effect size (η^2_p = 0.46), with task-congruent items rated lower than task-incongruent items. Domain also exhibited a significant main effect, *F*(5, 485) = 25.2, *p* < .001,η^2_p = 0.21, although specific mean scores for different domains are not provided in this summary. Additionally, there was a significant interaction between congruence and domain, *F*(5, 485) = 7.36, *p* < .001, η^2_p = 0.07.

For older adults, baseline MIQ ratings for each domain were analyzed with Welch's *t* tests. Here again, a Bonferroni correction with a cutoff score of α = 0.0042 was applied.

Significant differences between task-congruent and incongruent MIQ ratings were found in the domains of episodic memory ($p < .001$), working memory ($p = 0.002$), and processing speed ($p < .001$). No significant differences were detected in any of the other domains. Baseline ratings by domain can be found in **Table 6**.

In the post-test MIQ assessment of older adults (see **Table 7** for ratings), task-congruent items were rated significantly lower than task-incongruent items in the domains of episodic memory, semantic memory, working memory, and in processing speed (all *p'*s < .001). There were no significant differences in any of the other domains. Older adults' pre-assessment and post-assessment task-congruent and incongruent ratings can be found in **Table 6** and **Table 7** respectively.

	Younger Adults		
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**	***p***
Total *N* (*n* female)	98 (56)	98 (56)	
MIQ Ratings (*SD*)			
Global Score	0.37 (0.84)	0.57 (0.81)	< .001
Episodic Memory	0.33 (1.11)	0.81 (1.15)	0.001
Semantic Memory	0.91 (1.00)	0.55 (1.27)	0.02
Working Memory	-0.16 (1.46)	0.72 (1.38)	< .001
Executive Function (Stroop)	0.18 (1.37)	0.55 (1.14)	0.02
Executive Function (Rule ID)	0.55 (1.14)	0.18 (1.37)	0.02
Processing Speed	0.42 (1.16)	0.62 (1.21)	0.07

Table 4. Younger adults' pre-cognitive battery MIQ ratings for task-congruent vs. task-incongruent MIQ items. Standard deviations included in parentheses. Global score reflects the average of MIQ ratings across domains.

	Younger Adults		
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**	***p***
Total *N* (*n* females)	98 (56)	98 (56)	
MIQ Ratings (*SD*)			
Global Score	0.49 (0.85)	0.60 (0.88)	0.01
Episodic Memory	0.46 (1.09)	0.70 (1.24)	0.04
Semantic Memory	0.90 (1.03)	0.52 (1.38)	0.01
Working Memory	0.06 (1.38)	0.77 (1.31)	< .001
Executive Function (Stroop)	0.39 (1.27)	0.69 (1.10)	0.02
Executive Function (Rule ID)	0.69 (1.10)	0.39 (1.27)	0.02
Processing Speed	0.46 (1.28)	0.54 (1.27)	0.40

Table 5. Younger adults' post-cognitive battery MIQ ratings for task-congruent vs. task-incongruent MIQ items. Standard deviations are included in parentheses. The global score reflects the average of MIQ ratings across domains.

	Older Adults		
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**	***p***
Total *N* (*n* females)	97 (56)	97 (56)	
MIQ Ratings (*SD*)			
Global Score	0.24 (0.58)	0.58 (0.61)	< .001
Episodic Memory	-0.08 (0.76)	0.49 (1.0)	< .001
Semantic Memory	0.82 (0.88)	1.10 (0.88)	0.02
Working Memory	0.16 (1.01)	0.60 (0.93)	< .001
Executive Function (Stroop)	0.33 (1.04)	0.31 (0.88)	0.89
Executive Function (Rule ID)	0.31 (0.88)	0.33 (1.04)	0.89
Processing Speed	-0.10 (0.74)	0.62 (0.97)	< .001

Table 6. Older adults' pre- cognitive battery MIQ ratings for task-congruent vs. task-incongruent MIQ items. Standard deviations included in parentheses. Global score reflects the average of MIQ ratings across domains.

	Older Adults		
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**	***p***
Total *N* (*n* females)	97 (56)	97 (56)	
MIQ Ratings (*SD*)			
Global Score	0.18 (0.66)	0.51 (0.64)	<0.001
Episodic Memory	0.07 (0.81)	0.43 (0.92)	<0.001
Semantic Memory	0.62 (0.78)	1.06 (1.00)	<0.001
Working Memory	-0.13 (1.12)	0.45 (0.99)	<0.001
Executive Function (Stroop)	0.30 (1.08)	0.31 (0.89)	0.92
Executive Function (Rule ID)	0.31(0.89)	0.30 (1.08)	0.92
Processing Speed	-0.11 (0.80)	0.52 (1.04)	<0.001

Table 7. Older adults' post-cognitive battery MIQ ratings for task-congruent vs. task-incongruent MIQ items. Standard deviations are included in parentheses. Global score reflects the average of MIQ ratings across domains.

Effects of Task Congruence and Time Within Age Groups on Metacognitive Accuracy

Younger Adults. A 2 (congruence: task-congruent vs. incongruent) x 2 (time: pre- and post-task) ANOVA was conducted to examine differences in younger adults' task-congruent and task-incongruent metacognitive accuracy scores at two time points: pre- and post-cognitive battery assessment. The ANOVA indicated no significant main effect of time, *F*(1, 97) = 1.03, *p* = 0.31. However, a main effect of congruence was found, *F*(1, 97) = 18.7, *p* < .001, $\eta^2 p = 0.16$. Follow-up paired *t* tests revealed that younger adults’ accuracy scores, averaged across baseline scores and post-assessment scores, on the task-congruent items are significantly lower (*M* = 0.32; *SD* = 1.04) than on the task-incongruent items (*M* = 0.51; *SD* = 1.14).-To determine the accuracy differences, the mean absolute deviation (MAD) for both the task-congruent and task-incongruent items was computed. The aggregate MAD of the task-congruent items was lower, indicating higher metacognitive accuracy compared to the task-incongruent items. The interaction between congruence and time was not significant, *F*(1, 97) = 0.17, *p* = .68, indicating that task experience does not differentially affect younger adults’ metacognitive accuracy depending on how congruent those beliefs are with the task. Younger adults’ pre-assessment and post-assessment task-congruent and incongruent metacognitive accuracy scores can be found in **Table 9** and **Table 10** respectively. For comparison's sake, pre- and post-cognitive task metacognitive accuracy scores collapsed across congruency for each age group can be found in **Table 8**.

Older Adults. A 2 (congruence: task-congruent vs. incongruent) x 2 (time: pre- and post-task) ANOVA was conducted to examine differences in older adults’ task-congruent and task-incongruent metacognitive accuracy scores at two time points: pre- and post-cognitive battery assessment. The ANOVA indicated a significant main effect of congruence, *F*(1, 96) = 140.2, *p* < .001, $\eta^2_p = .59$, and a significant main effect of time, *F*(1, 96) = 26.1, *p* < .001, $\eta^2_p = 0.21$. There was also a significant interaction between congruence and time, *F*(1, 96) = 25.8, *p* < .001, $\eta^2_p = .21$, suggesting that the effect of time on metacognitive accuracy is influenced by whether the assessments made are task-congruent or task-incongruent.-To decompose this interaction, follow-up paired t-tests were conducted and revealed that older adults' accuracy scores were significantly lower on task-congruent MIQ items after taking the cognitive battery (i.e., at time 2) (*M* = -0.01; *SD* = 0.85) compared to before taking the battery (*M* = 0.26; *SD* = 0.87); t(96) = 6.51, *p* < .001, *d* = 0.66. The MAD for these mean scores indicates a higher metacognitive accuracy on these items post-test than at baseline. For incongruent-MIQ items, older adults' accuracy scores did not significantly differ from pre-task (*M* = 0.60; *SD* = 0.90) to post-task (*M* = 0.53; *SD* = 0.91), *t*(96) = 1.95, *p* = .054, *d* = 0.20. Older adults' pre-assessment and post-assessment task-congruent and incongruent ratings can be found in **Table 11** and **Table 12** respectively.

	Age Group			
	Younger Adults		**Older Adults**	
Timepoint	**Pre-Test**	**Post-Test**	**Pre-Test**	**Post-Test**
Total *N* (*n* females)	98 (56)	98 (56)	97 (56)	97 (56)
Metacognitive Accuracy (*SD*)				
Global Score	0.49 (1.60)	0.50 (1.59)	0.44 (1.36)	0.25 (1.35)
Episodic Memory	0.63(1.42)	0.64 (1.44)	0.16 (1.13)	0.21(1.14)
Semantic Memory	0.63 (1.15)	0.40 (0.59)	1.10 (1.32)	0.38 (0.68)
Working Memory	0.28 (1.50)	0.42 (1.42)	0.32 (1.02)	0.10 (1.01)
Executive Function	0.34 (1.26)	0.51 (1.24)	0.35 (1.13)	0.34 (1.12)
Processing Speed	0.56 (1.58)	0.54 (1.63)	0.29 (1.25)	0.24 (1.26)

Table 8. Pre- and post-cognitive battery metacognitive accuracy scores–collapsed across task congruency–for younger and older adults. Standard deviations are included in parentheses. Global score reflects the average of accuracy scores across domains.

	Younger Adults	
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**
Total *N* (*n* females)	98 (56)	98 (56)
Metacognitive Accuracy (*SD*)		
Global Score	0.29 (1.13)	0.50 (1.16)
Episodic Memory	0.39 (1.52)	0.87 (1.64)
Semantic Memory	0.81 (1.25)	0.45 (1.45)
Working Memory	-0.16 (1.66)	0.73 (1.78)
Executive Function (Stroop)	0.12 (1.73)	0.56 (1.36)
Executive Function (Rule ID)	0.56 (1.36)	0.12 (1.73)
Processing Speed	0.46 (1.58)	0.66 (1.76)

Table 9. Younger adults' pre-cognitive battery metacognitive accuracy scores for task-congruent vs. task-incongruent MIQ items. Standard deviations are included in parentheses. Global score reflects the average of accuracy scores across domains.

	Younger Adults	
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**
Total *N* (*n* females)	98 (56)	98 (56)
Metacognitive Accuracy (*SD*)		
Global Score	0.35 (1.03)	0.53 (1.17)
Episodic Memory	0.52 (1.44)	0.76 (1.65)
Semantic Memory	0.38 (1.42)	0.42 (1.59)
Working Memory	0.06 (1.62)	0.77 (1.65)
Executive Function (Stroop)	0.32 (1.58)	0.70 (1.45)
Executive Function (Rule ID)	0.70 (1.45)	0.32 (1.58)
Processing Speed	0.50 (1.68)	0.58 (1.71)

Table 10. Younger adults’ post-cognitive battery metacognitive accuracy scores for task-congruent vs. task-incongruent MIQ items. Standard deviations are included in parentheses. Global score reflects the average of accuracy scores across domains.

	Older Adults	
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**
Total *N* (*n* females)	97 (56)	97 (56)
Metacognitive Accuracy (*SD*)		
Global Score	0.26 (0.87)	0.60 (0.90)
Episodic Memory	-0.13 (1.09)	0.45 (1.39)
Semantic Memory	0.96 (1.46)	1.24 (1.37)
Working Memory	-0.10 (1.20)	0.54 (1.15)
Executive Function (Stroop)	0.36 (1.44)	0.34 (1.24)
Executive Function (Rule ID)	0.34 (1.24)	0.36 (1.44)
Processing Speed	-0.07 (1.25)	0.65 (1.41)

Table 11. Older adults' pre-cognitive battery metacognitive accuracy scores for task-congruent vs. task-incongruent MIQ items. Standard deviations included in parentheses. Global score reflects the average of accuracy scores across domains.

	Older Adults	
Item Type (MIQ)	**Task-Congruent Items**	**Task-Incongruent Items**
Total *N* (*n* females)	97 (56)	97 (56)
Metacognitive Accuracy (*SD*)		
Global Score	-0.01 (0.85)	0.53 (0.91)
Episodic Memory	0.03 (1.10)	0.39 (1.35)
Semantic Memory	-0.44 (1.02)	1.20 (1.39)
Working Memory	-0.19 (1.19)	0.39 (1.17)
Executive Function (Stroop)	0.33 (1.54)	0.34 (1.27)
Executive Function (Rule ID)	0.34 (1.27)	0.33 (1.54)
Processing Speed	-0.08 (1.22)	0.55 (1.47)

Table 12. Older adults' post-cognitive battery metacognitive accuracy scores for task-congruent vs. task-incongruent MIQ items. Standard deviations are included in parentheses. Global score reflects the average of accuracy scores across domains.

Discussion

The present study had two main aims. First, it aimed to assess the validity of our approach for measuring global metacognition, specifically by examining the extent to which global metacognitive beliefs differ between task-congruent and task-incongruent assessments.

The second aim was to examine how task exposure and age interact to affect global metacognition.

One of the main findings of this study was that younger adults exhibited task-induced belief updating but only for global metacognitive beliefs most relevant to our cognitive battery—a finding obfuscated when analyzing the data as Arar et al. (in prep) did. Initial work from Arar et al. found that younger adults' metacognitive belief ratings and accuracy remained consistent from pre- to post-cognitive battery, which suggested that perhaps younger adults were not as metacognitively savvy. Results from the present study indicate the opposite and offer a more complete and nuanced interpretation. Our findings would suggest that younger adults were able to selectively and effectively use task experience to inform beliefs most relevant to task experience, whereas task-incongruent beliefs were informed by cues unrelated to their recent task-experience (which cues exactly are for future work to address).

Older adults, however, did not demonstrate this selectivity—at least, with regard to MIQ beliefs. No effect of congruence was found on older adults' belief updating from Time 1 to Time 2. Instead, older adults rated themselves lower on all MIQ items at Time 2 compared to Time 1, which may be evidence for stereotype threat effects. Future work focused on disentangling transient stereotype threat effects from genuine, long-lasting updates to older adults' self-concepts is also needed.

Overall, this disparity in the interaction of time and congruence in the MIQ ratings observed in younger adults, but absent in older adults, suggests that younger adults may possess a more sophisticated metacognitive ability to adjust their self-assessments based on the relevance and direct experience of specific cognitive tasks. In other words, younger adults have a greater capacity to selectively update their global metacognitive beliefs, applying task-specific insights to adjust their beliefs in a targeted manner compared to older adults.

The effects of congruence and domain on global metacognitive ratings offer a nuanced understanding of how younger and older adults perceive their cognitive abilities. For younger adults, significant differences emerged when their ratings were analyzed by cognitive domain. Notably, they rated their episodic and working memory abilities lower on task-congruent items compared to task-incongruent items. Such differences within specific domains may stem from task difficulty, as the working memory task was generally the most challenging. Alternatively, differences could be attributed to the phrasing of the items on the MIQ. For example, the task-congruent assessment for working memory involves "rehearsing phone numbers or directions," potentially reflecting a more demanding task involving both digit-span and directional abilities, leading to underestimation. Conversely, the task-incongruent assessment focuses on "multi-tasking (e.g., cooking while talking)," a seemingly more common everyday task, which may instill higher confidence.

Interestingly, this pattern was not consistent across all cognitive domains. No significant differences were found in executive function, semantic memory, and processing speed, raising questions about the influence of the degree of task-congruency on metacognitive judgments. For instance, task-congruent items probing executive function, semantic memory, and processing speed closely mirrored the tasks presented on the battery. In contrast, task-congruent items for episodic memory may be less directly related to the cognitive battery tasks. For instance, task-congruent items probing executive function, semantic memory, and processing speed closely mirrored the tasks presented on the battery. For example, for semantic memory, the task-congruent questionnaire item “identifying common objects” (MIQ) almost exactly represents the semantic memory cognitive battery task– Object Identification and Naming Task. Similarly, for the processing speed task on the battery, the participants must perform the Fast Comparison Task with utmost speed. This task is exactly reflected in the task-congruent questionnaire item asking for their self-assessment on “Performing speeded tasks.” (MIQ). In contrast, task-congruent items for episodic memory may be less directly related to the cognitive battery tasks.

For older adults, baseline MIQ ratings also showed significant effects of congruence, with task-congruent items rated lower than task-incongruent items, particularly in episodic, memory, working memory and processing speed domains. This pattern persisted into post-test assessments, where task-congruent items in episodic memory, working memory, and processing speed domains were rated lower than incongruent items. Additionally, there was a significantly higher MIQ rating for task-incongruent items compared to task-congruent items in the semantic memory domain post-test. These findings suggest that older adults may adjust their metacognitive evaluations based on task relevance, although the magnitude of these adjustments may vary across different cognitive domains. This suggests a complex belief updating process that is influenced by an interaction among age, cognitive domain, and task congruence.

This research underscores why developing appropriate measurement techniques for global metacognition is critical and advances our understanding of how factors—like task exposure and age—come to bear on global metacognitive updating and formation. Moreover, these results have potential applications in cognitive training. The different emerging patterns in metacognitive judgment adjustment between younger and older adults noted above offer insights for designing targeted interventions aimed at enhancing metacognitive accuracy and cognitive functioning across the lifespan. However, the proposed method has limitations. It primarily focuses on metacognitive accuracy within specific cognitive domains, potentially limiting its applicability to a broader range of cognitive abilities, such as perceptual skills. Additionally, separating items based on task-congruency and considering only one metacognitive judgment per domain as task-congruent may provide a limited view of participants' overall self-assessment of their cognitive abilities. Further research is needed to explore the factors involved in global metacognitive belief formation and updating over time, across diverse populations and cognitive domains.

References

Arar, T. (2021). Measuring Global and Local Metacognition: Effects of Aging and Cognitive Task Experience [Master's thesis, University of Chicago].

Bertrand, J. M., Moulin, C. J., & Souchay, C. (2017). Short-term memory predictions across the lifespan: Monitoring span before and after conducting a task. Memory, 25, 607–618.https://doi.org/10.1080/09658211.2016.1200625

Brosnan, S. F., Wilson, B. J., & Beran, M. J. (2012). Old World monkeys are more similar to humans than New World monkeys when playing a coordination game. Proceedings of the Royal Society B: Biological Sciences, 279(1733), 1522-1530.https://doi.org/10.1098/rspb.2011.1781

David, A. S., Bedford, N., Wiffen, B., & Gilleen, J. (2012). Failures of metacognition and lack of insight in neuropsychiatric disorders. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1594), 1379-1390.https://doi.org/10.1098/rstb.2012.0002

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. Psychology and Aging, 26, 631–635.https://doi.org/10.1037/a0022852

Hargis, M. B., & Castel, A. D. (2018). Improving Medication Understanding and Adherence Using Principles of Memory and Metacognition. Policy Insights from the Behavioral and Brain Sciences, 5(2), 147-154.https://doi.org/10.1177/2372732218781643

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. Current Directions in Psychological Science, 20(3), 167-173.https://doi.org/10.1177/0963721411409026

Hertzog, C., Dunlosky, J., Powell-Moman, A., & Kidder, D. P. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? Psychology and Aging, 17, 209–225. https://doi.org/10.1037/0882-7974.17.2.209

McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacognitive monitoring: The role of consequences, task experience, and prior knowledge. Experimental Aging Research, 43(3), 233-256.https://doi.org/10.1080/0361073X.2017.1298956

McWilliams, A., Bibby, H., Steinbeis, N., David, A. S., & Fleming, S. M. (2022). Age-Related Decreases in Global Metacognition Are Independent of Local Metacognition and Task Performance. [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/nmhxv

Robinson, E. A., Hertzog, C., & Dunlosky, J. (2006). Aging, encoding fluency, and metacognitive monitoring. Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition, 13, 458–478.https://doi.org/10.1080/13825580600572983

Rouault, M., & Fleming, S. M. (2020). Formation of global self-beliefs in the human brain. Proceedings of the National Academy of Sciences, 117, 27268-27276.https://doi.org/10.1073/pnas.2003094117

Rouault, M., Seow, T., Gillan, C. M., & Fleming, S. M. (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biological Psychiatry, 84(6), 443–451.https://doi.org/10.1016/j.biopsych.2017.12.017

Salthouse, T. A. (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society, 16(5), 754-760. https://doi.org/10.1017/S1355617710000706

Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. Consciousness and Cognition, 16, 769–784.https://doi.org/10.1016/j.concog.2006.11.002

Starkstein, S. E., Jorge, R., Mizahi, R., Adrian, J., & Robinson, R. G. (2007). Insight and danger in Alzheimer's disease. European Journal of Neurology, 14(4), 455–460. https://doi.org/10.1111/j.1468-1331.2007.01745.x

Wong, J. T., & Gallo, D. A. (2019). Activating aging stereotypes increases source recollection confusions in older adults: effect at encoding but not retrieval. The Journals of Gerontology: Series B, 74, 633-641. https://doi.org/10.1093/geronb/gbx103