UCLA

Title

Knowing more than we know: metacognition, semantic fluency, and originality in younger and older adults.

Authors

Murphy, Dillon

Castel, Alan

Peer reviewed

Knowing More than We Know: Metacognition, Semantic Fluency, and Originality in Younger and Older Adults

Dillon H. Murphy,

Alan D. Castel

University of California Los Angeles, Los Angeles, CA, USA

Abstract

We examined age-related similarities and differences in people's metacognitive awareness of retrieval from semantic long-term memory as well as the originality of their responses. Participants completed several semantic fluency tasks, and before recalling items, made metacognitive predictions of their performance. Additionally, after retrieval, participants made metacognitive evaluations of the originality of their responses. Results revealed that both younger (M_{age} = 24.49) and older adults (M_{age} = 68.31) were underconfident in their performance, despite some metacognitive awareness of their ability to retrieve information from semantic memory. Younger and older adults became more metacognitively aware of their abilities with task experience, but there were no significant differences in participants' metacognitive predictions and postdictions, although older adults believed that they were less original than younger adults. These findings revealed a "skilled and unaware" effect whereby participants were underconfident on the first trial and became less underconfident on later trials. These patterns may fit with a broader literature that has found a lack of adult age differences in metacognition for verbal skills but shows that older adults may believe that their access to original verbal knowledge may decline in older age.

Keywords

semantic fluency; originality; metacognition; aging; retrieval

Semantic fluency tasks assess the retrieval ability of knowledge from long-term memory by having participants generate items from a given category (e.g., animals; Gruenewald & Lockhead, 1980). Fluency tasks can be informative in terms of differences in both cognitive and neuropsychological function, and fluency performance deficits have been shown in older adults (Murphy & Castel, 2021; Rodríguez-Aranda & Martinussen, 2006; Troyer et al., 1997), patients with frontal and temporal lobe deficits (Henry & Crawford, 2004), patients with Alzheimer's, Parkinson's, and Huntington's disease (Troster et al., 1998), as well as individuals low in intelligence and working memory capacity (Rosen & Engle, 1997).

While the goal of fluency tasks is to generate as many items as possible, researchers have also examined the qualitative aspects of participants' retrieval, specifically, the originality of retrieved exemplars. The process of generating original ideas has been extensively examined in various tasks (see Boot et al., 2017; Sowden et al., 2015, for reviews) and in semantic fluency tasks, originality focuses on the relative uniqueness of retrieved items, often assessed via retrieval infrequency (see Dumas & Dunbar, 2014; Hocevar, 1979; Silvia, 2008, 2011 for various methods). Generally, fluency and originality scores are positively related such that people who recall more items overall tend to be more unique in their responses (e.g., Dumas & Dunbar, 2014; Silvia, 2008; but see Benedek et al., 2006; Hocevar, 1980). As people age, older adults may experience some declines in verbal fluency and originality (Murphy & Castel, 2021; Rodríguez-Aranda & Martinussen, 2006; Troyer et al., 1997) but it is unclear to what degree younger and older adults are aware of these potential age-related differences.

Despite older adults having good semantic memory and often better vocabulary knowledge than younger adults (Ben-David & Erel, 2015; Kavé & Halamish, 2015), younger adults sometimes perform better on fluency tasks (e.g., Rodríguez-Aranda & Martinussen, 2006; Troyer et al., 1997) and can also be more original in their responses than older adults. However, some aspects of verbal fluency may be well-maintained, to a certain age, in healthy older adults (Murphy & Castel, 2021). This exemplifies some of the variety of cognitive changes accompanying healthy aging (cf., Hess, 2005; Park & Festini, 2017; Salthouse, 2010), and evidence suggests that older adults are aware of many of these deficits. For example, Kavé and Halamish (2015) examined age-related differences in the ability to judge one's accuracy on a multiple-choice test of vocabulary and also estimate their overall performance. Older adults were better at assessing the accuracy of their knowledge compared with younger adults, as evidenced by their better calibration (match between prediction and performance; see also Hertzog et al., 2010).

The awareness and understanding of one's memory abilities is known as metacognition (Dunlosky et al., 2016; Nelson, 1996; Nelson & Narens, 1990). To assess the accuracy of metacognition, researchers often ask participants to judge how likely they are to later remember information, a form of metacognitive monitoring (see Rhodes, 2016 for a review). Although memory performance generally declines with age, the accuracy of metacognitive monitoring is generally intact in older age (Hertzog & Dunlosky, 2011). However, research on metacognitive monitoring has largely focused on episodic memory tasks (e.g., Halamish et al., 2011; Hertzog et al., 2002; Stine-Morrow et al., 2006) rather than retrieval from semantic memory (but see Rhodes & Castel, 2008), and has not addressed how younger and older adults assess the originality of information that is retrieved from semantic memory.

Whether completing semantic (e.g., Kavé & Halamish, 2015) or episodic memory tasks (e.g., Hertzog & Dunlosky, 2011), younger and older adults are often similarly accurate in their metamemory abilities at the item level (specific exemplars), but older adults are more accurate at the global level (estimates of overall performance; but see Dodson et al., 2007; Souchay et al., 2007). However, despite some initial overconfidence (especially in older adults), younger and older adults tend to become more metacognitively accurate with increased task experience (e.g., McGillivray & Castel, 2011), and may even experience

a form of "underconfidence with practice" when learning the same material in multiple sessions (Tauber & Rhodes, 2012). Furthermore, researchers have also investigated younger and older adults' ability to monitor performance that has already occurred. When asked to make postdictive judgments about the accuracy of their memory, older adults perform similarly to younger adults (Baker et al., 2010), again indicating older adults' metacognitive awareness of aging deficits.

While metacognitive monitoring has largely focused on overall memory performance, little work has investigated participants' awareness of the qualitative aspects of their recall. Some work has examined how people evaluate the originality of ideas generated by others (see Runco & Charles, 1993) but a recent study introduced a metacognitive judgment of originality and demonstrated that younger participants systematically underestimate the originality of their ideas (Sidi et al., 2020). Thus, similar to metacognitive judgments in episodic memory tasks, originality judgments can be prone to biases. Specifically, in contrast to the ubiquitous observation that participants are overconfident in memory task performance (see Dunlosky & Metcalfe, 2008; Koriat, 2016, for reviews), people (especially older adults) may systematically underestimate their originality but become more accurate with increased task experience. Rather than a prediction of whether or not information can be recalled later (an objective performance prediction), originality judgments rely on a comparison to others which involves greater uncertainty (see Moore & Healy, 2008).

The Current Study

The goal of the current study was to gain a better understanding of people's metacognitive awareness of retrieval from semantic long-term memory as well as the originality of their responses. Participants completed semantic fluency tasks but before recalling items, made metacognitive predictions of their performance. Additionally, after retrieval, participants made metacognitive evaluations of the originality of their responses. In terms of overall fluency performance, we expected participants to be overconfident but to become better calibrated with increased task experience. We also expected participants to become more aware of their originality with increased task experience. In Experiment 1, we tested younger adults using common verbal fluency tasks and added novel measures that examine metacognitive predictions of both fluency and originality. In Experiment 2, we examined and compared these measures in younger and older adults.

Since one of our main research objectives was to examine how younger and older adults assess the originality of information that is retrieved from semantic memory, we wanted to ensure that participants could view their responses while assessing originality so that this measure was not biased/influenced by interference and retrieval failure which could be more pronounced in older adults and influence judgments (e.g., Coane & Umanath, 2019; Murphy & Castel, 2022; Murphy Castel, 2022b; Witherby et al., 2019). While some prior work has examined metacognitive monitoring without providing access to responses, we felt it was important for participants to have access to the responses to allow for a comparison of originality that is not influenced by potential biases in memory.

Experiment 1

In Experiment 1, we sought to establish a paradigm to test younger adults' semantic fluency and originality by using standard tasks of semantic fluency (generating exemplars of categories from semantic memory) and then having participants rate the originality of their responses. Younger adults completed two trials of the semantic fluency task and made global predictions of performance before each trial. Additionally, after each trial, participants provided a judgment of the originality of their responses. Since we were interested in participants' global assessment of originality, not just the number of original responses, we asked participants to provide a general assessment of their performance in terms of originality. We expected younger adults to be initially overconfident in their fluency predictions but to become more metacognitively accurate with increased task experience. Similarly, we expected younger adults to become more accurate in their originality judgments after gaining task experience.

Method

Participants.—Participants were 103 younger adults (age range: 18-27, M_{age} = 19.91, SD_{age} = 1.66; in terms of the highest level of education achieved, 1 some High School, 24 High School Graduate, 62 some college but no degree, 11 Associates degree, 5 Bachelor's degree) recruited from the University of California Los Angeles (UCLA) Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were required to be fluent in English to participate. Participants were excluded from analysis if they admitted to cheating (e.g., looking up answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in zero exclusions. A sensitivity analysis based on the observed data indicated that for a paired samples t-test, assuming alpha = .05, power = .80, for a two-tailed test, the smallest effect (comparisons between trials) the design could reliably detect is d = .20.

Materials and Procedure.—Participants were told that they would be asked to retrieve as many items from a category as they could in 5 minutes. After being told the category with which they would retrieve items, participants predicted how many items they would retrieve. Participants were informed that they could retrieve items in any order that they wished but that they should keep trying to retrieve items throughout the entire 5-minute period (they typed their responses into an on-screen text box and responses remained on screen after they were typed). After the retrieval period, participants were asked to rate the uniqueness of their responses on a scale from 0 (not unique) to 10 (very unique). For example, participants were asked, "How original/unique do you think the animals you retrieved are? Please answer on a scale from 0 (not original/unique) to 10 (very original/unique)". Participants then repeated this procedure for a second category of items. Categories included "animals" and "foods you can buy at the supermarket." These categories were selected because they have easily defined correct responses and have been found to be related to other fluency measures (e.g., Unsworth et al., 2011). The order of the categories was counterbalanced to control for any scaling issues or category effects.

Scoring.: For the category "animals," any attempt to retrieve a type of living animal, real or fake, was counted as correct. Repetitions (retrieving the same item more than once), semantic repetitions (e.g., deer and buck), non-exemplars (e.g., reptiles), and submissions that were not animals were counted as incorrect. Within the data coding process, participants' incorrect responses due to spelling errors or typos were counted if the intent of the response was clear (e.g., the response "baboon" followed by the response "arangatang" was interpreted as "orangutan"). For the category "foods you can buy at the supermarket," any attempt to retrieve an edible food item was counted as correct. Again, repetitions, semantic repetitions, non-exemplars, and submissions that were not foods were counted as incorrect and spelling errors were counted if the intent of the response was clear.

Four scores were computed for each participant in each category: (a) the number of items generated, excluding errors and repetitions (fluency), (b) the number of participants' unique responses (originality), (c) the difference between participants' predictions of performance and the number of items generated (i.e., *calibration*; Rhodes, 2016), and (d) their originality rating. Based on the recommendations of Milgram and Milgram (1976), responses generated by about 5% or less of the sample (the number of items per participant that were generated 5 times or less out of the 103 participants) were considered original. Examples of items scored as common and original are shown in Table 1.

Results

Correlations between fluency and originality performance and metacognitive predictions/ postdictions on each trial are shown in Table 2. In addition to descriptive and inferential statistics, to provide evidence for the strength of each effect, we also computed the Bayes Factor (BF; a ratio of the marginal likelihood of the null model and a model suggesting group differences). The data are reported as either in favor of the null hypothesis (which would be supported by a large BF_{01}) or the alternative hypothesis (which would be supported by a large BF_{10}; see Kass & Raftery, 1995 for more information). To determine whether retrieval fluency improved with task experience, a paired samples *t*-test indicated that the number of items retrieved on the first trial ($M = 51.14$, $SD = 16.72$) was similar to the second trial ($M = 51.90$, $SD = 14.69$), [$t(102) = .58$, $p = .566$, $d = .06$, $BF_{01} = 7.79$]. To determine whether fluency predictions changed with task experience, a paired samples *t*-test indicated that the number of items expected to be retrieved on the first trial ($M = 24.96$, $SD = 19.83$) was less than the second trial ($M = 31.86$, $SD = 18.48$), [$t(102) = 3.88$, $p < .001$, $d = .38$, $BF_{10} > 100$].

To examine the accuracy of participants' predictions of their performance (see Figure 1), we calculated *calibration* (i.e., the difference between the number of items participants expected to retrieve and the number of items they actually retrieved). Across trials, a one-sample *t*-test confirmed that participants were underconfident [$t(102) = -14.39$, $p < .001$, $d = -1.42$, $BF_{10} > 100$] and participants were underconfident on both trials [1st: $t(102) = -12.85$, $p < .001$, $d = -1.27$, $BF_{10} > 100$; 2nd: $t(102) = -11.02$, $p < .001$, $d = -1.09$, $BF_{10} > 100$]. A paired samples *t*-test indicated that participants became better calibrated with increased task experience [$t(102) = 2.86$, $p = .005$, $d = .28$, $BF_{10} = 5.06$]. Specifically, calibration on the

second trial ($M = -20.04$, $SD = 18.45$) was better than the first trial ($M = -26.17$, $SD = 20.68$).

To examine participants' originality as a function of task experience (see Figure 2a), a paired samples t-test revealed that the number of original items retrieved on the second trial ($M = 7.92$, $SD = 8.46$) was greater than the first trial ($M = 6.24$, $SD = 6.21$), [$t(102) = 2.07$, $p = .041$, $d = .20$, $BF_{10} = .84$]; since Bayes Factor suggests that there is no convincing evidence for this effect, we do not further interpret this finding (which was not replicated in Experiment 2). To examine participants' postdictions as a function of task experience (which were on a Likert scale and should not be compared to the number of original exemplars retrieved; see Figure 2b), a paired samples t-test revealed that postdictions on the first trial ($M = 4.65$, $SD = 2.55$) were similar to the second trial ($M = 4.97$, $SD = 2.63$), [$t(102) = 1.43$, $p = .155$, $d = .14$, $BF_{01} = 3.40$].

Discussion

In Experiment 1, participants retrieved a similar number of items on each trial but were greatly underconfident in their performance. However, participants were more confident on the second trial and subsequently became better calibrated on the second task. Furthermore, on both trials, participants' predictions and fluency performance were strongly positively correlated indicating some metacognitive awareness of performance. In terms of originality, participants became more original with increased task experience, but their judgments of their originality were similar on each trial. However, participants' originality and their judgments of originality were unrelated on the first trial but positively correlated on the second trial, indicating an increased metacognitive awareness of originality with increased task experience.

Experiment 2

In Experiment 1, younger adults were underconfident in their fluency performance but became better calibrated on the second trial and were more accurate in their originality postdictions on the second trial. In Experiment 2, we investigated whether these effects extend to older adults. While some prior work has found age-related deficits in fluency and originality (e.g., Rodríguez-Aranda & Martinussen, 2006; Troyer et al., 1997), these effects may be somewhat minimal in fairly healthy older adults under the age of 70 (see Murphy & Castel, 2021), but older adults may be acutely aware of any potential impairments. We also added a third trial to further elucidate whether people become better metacognitively tuned after gaining task experience. We expected older adults to be more metacognitively accurate than younger adults and for both younger and older adults to become better calibrated with increased task experience. In terms of originality, we expected older adults to be less metacognitively accurate in their originality judgments, but for both younger and older adults to become more metacognitively accurate in their originality postdictions with increased task experience.

Method

Participants.—After exclusions, participants were 119 younger adults (age range: 19-29, M_{age} = 24.49, SD_{age} = 2.10; 15 High School Graduate, 26 some college but no degree, 6 Associates degree, 59 Bachelor's degree, 13 Graduate degree (Masters, Doctorate, etc.)) and 107 older adults (age range: 64-79, M_{age} = 68.31, SD_{age} = 3.57; 2 some High School, 11 High School Graduate, 22 some college but no degree, 11 Associates degree, 32 Bachelor's degree, 29 Graduate degree) recruited from Amazon's Mechanical Turk, a Web site that allows users to complete small tasks for pay. Participants received \$2.50 for completing the experiment, which took approximately 20 minutes. All participants were required to be in the United States to participate. Participants were required to be fluent in English to participate. Participants were excluded from analysis if they admitted to cheating (e.g., looking up answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in two exclusions from the younger adult group and no exclusions from the older adult group. A sensitivity analysis based on the observed data indicated that for a 2 (age: young, old) x 3 (trial) mixed ANOVA, assuming alpha = .05, power = .80, with a high correlation between repeated measures (r = .75), the smallest effect (age-related differences) the design could reliably detect is η^2 = .03.

Materials and Procedure.—The task in Experiment 2 was similar to Experiment 1 but included a third category: "occupations." Similar to Experiment 1, we counterbalanced the order of the categories to control for any scaling issues or category effects. At the end of the task, participants were asked to estimate the number of items they could generate for each of the three categories if they were to complete the task again. Additionally, for each category, we asked participants how many items they would expect the average younger adult (aged 18-30) to retrieve and how many items they would expect the average older adult (aged 65-85) to retrieve.

Scoring.: Scoring for the category "animals" and "foods you can buy at the supermarket" was the same as in Experiment 1. For the category "occupations," any attempt to retrieve a legitimate job or profession was counted as correct (see Table 1 for examples). Again, repetitions, semantic repetitions, non-exemplars, and submissions that were not occupations were counted as incorrect and spelling errors were counted if the intent of the response was clear. Finally, original items were responses generated by about 5% or less of the sample (the number of items per participant that were generated 11 times or less out of the 226 participants).

Results

Predictions and performance.—Correlations between fluency and originality performance, metacognitive predictions/postdictions on each trial, and age (as a continuous variable) are shown in Table 3. To examine whether fluency predictions changed with task experience, a 2 (age: young, old) x 3 (trial) mixed ANOVA did not reveal a main effect of trial [$F(2, 448)$ = 1.33, p = .267, η^2 = .01, BF_{01} = 17.32] such that participants predicted that they would retrieve a similar number of exemplars on the first trial (M = 27.04, SD = 24.19), second trial (M = 26.16, SD = 19.41), and third trial (M = 25.01, SD = 18.46). Additionally, results did not reveal a main effect of age [$F(1, 224)$ = .22, p = .641, η^2 < .01, BF_{01} = 4.83]

such that younger adults predicted that they would retrieve a similar number of exemplars ($M = 25.49$, $SD = 18.61$) on each trial as older adults ($M = 26.59$, $SD = 17.09$), and trial did not interact with age [$F(2, 448) = .20$, $p = .822$, $\eta^2 < .01$, $BF_{01} = 25.29$].

To examine whether retrieval fluency improved with task experience, a 2 (age: young, old) x 3 (trial) mixed ANOVA did not reveal a main effect of trial [$F(2, 448) = .76$, $p = .467$, $\eta^2 < .01$, $BF_{01} = 29.55$] such that participants retrieved a similar number of exemplars on the first trial ($M = 29.83$, $SD = 13.12$), second trial ($M = 30.59$, $SD = 13.69$), and third trial ($M = 30.45$, $SD = 15.11$). Additionally, results did not reveal a main effect of age [$F(1, 224) = .03$, $p = .870$, $\eta^2 < .01$, $BF_{01} = 4.15$] such that younger adults retrieved a similar number of exemplars ($M = 30.16$, $SD = 9.13$) on each trial as older adults ($M = 30.44$, $SD = 15.32$), and trial did not interact with age [$F(2, 448) = .59$, $p = .557$, $\eta^2 < .01$, $BF_{01} = 15.48$].

To examine the accuracy of participants' predictions of their performance (see Figure 3), we calculated younger and older adults' calibration. Across trials, a one-sample t-test confirmed that participants were underconfident [$t(225) = -3.70$, $p < .001$, $d = -.25$, $BF_{10} = 52.27$] and participants were underconfident on most trials [1st: $t(225) = -1.74$, $p = .083$, $d = -.12$, $BF_{01} = 3.03$; 2nd: $t(225) = -3.66$, $p < .001$, $d = -.24$, $BF_{10} = 46.19$; 3rd: $t(225) = -4.66$, $p < .001$, $d = -.31$, $BF_{10} > 100$] but a 2 (age: young, old) x 3 (trial) mixed ANOVA did not reveal a main effect of trial [$F(2, 448) = 2.47$, $p = .086$, $\eta2 = .01$, $BF_{01} = 6.15$], a main effect of age [$F(1, 224) = .37$, $p = .545$, $\eta^2 < .01$, $BF_{01} = 4.58$], and trial did not interact with age [$F(2, 448) = .35$, $p = .704$, $\eta^2 < .01$, $BF_{01} = 21.30$]. Thus, although both younger and older adults were still underconfident, they did not become better calibrated with increased task experience.

To examine participants' originality as a function of task experience (see Figure 4a), a 2 (age: young, old) x 3 (trial) mixed ANOVA did not reveal a main effect of trial [$F(2, 448) = .04$, $p = .965$, $\eta^2 < .01$, $BF_{01} = 56.10$] such that participants retrieved a similar number of original items on the first trial ($M = 6.82$, $SD = 6.47$), second trial ($M = 6.83$, $SD = 6.19$), and third trial ($M = 6.95$, $SD = 6.68$). Additionally, results did not reveal a main effect of age [$F(1, 224) = .23$, $p = .629$, $\eta^2 < .01$, $BF_{01} = 5.52$] such that younger adults were similarly original ($M = 7.03$, $SD = 6.10$) on each trial as older adults ($M = 6.69$, $SD = 4.14$). Additionally, trial did not interact with age [$F(2, 448) = 2.02$, $p = .134$, $\eta^2 = .01$, $BF_{01} = 4.70$].

To examine participants' originality postdictions as a function of task experience (see Figure 4b), a 2 (age: young, old) x 3 (trial) mixed ANOVA revealed a main effect of trial [$F(2, 448) = 4.01$, $p = .019$, $\eta^2 = .02$, $BF_{10} = .93$] such that participants believed that they were more original on the second trial ($M = 5.19$, $SD = 2.64$) than the first trial ($M = 4.84$, $SD = 2.71$), [$p_{bonf} = .045$, $d = .16$] and marginally more original on the third trial ($M = 5.18$, $SD = 2.58$) than the first trial [$p_{bonf} = .057$, $d = .16$] but there were no differences between the second and third trials [$p_{bonf} > .999$, $d < .01$]; since Bayes Factor did not support this main effect, we advise caution in the interpretation of this finding. Furthermore, results revealed a main effect of age [$F(1, 224) = 26.53$, $p < .001$, $\eta^2 = .11$, $BF_{10} > 100$] such that younger adults believed that they were more original ($M = 5.80$, $SD = 2.15$) on each trial than older adults ($M = 4.26$, $SD = 2.35$), but trial did not interact with age [$F(2, 448) = 1.09$, $p = .338$, $\eta^2 = .01$, $BF_{01} = 11.14$].

Post-task questionnaire.—To examine participants' post-task predictions of fluency for each category (predictions of performance if they were to try again), a 2 (age: young, old) x 3 (category) mixed ANOVA revealed a main effect of category [$F(2, 436) = 16.49$, $p < .001$, $\eta^2 = .07$, $BF_{10} > 100$] such that participants predicted that they would retrieve more foods ($M = 35.72$, $SD = 22.28$) than occupations ($M = 27.80$, $SD = 15.25$), [$p_{bonf} < .001$, $d = .58$] but not animals ($M = 31.65$, $SD = 27.32$), [$p_{bonf} = .061$, $d = .16$]; additionally, participants predicted that they would retrieve more animals than occupations [$p_{bonf} = .010$, $d = .20$]. However, results did not reveal a main effect of age [$F(1, 218) = .12$, $p = .729$, $\eta^2 < .01$, $BF_{01} = 5.19$] such that younger adults predicted that they would retrieve a similar number of exemplars ($M = 31.32$, $SD = 15.85$) on each category as older adults ($M = 32.14$, $SD = 21.52$), and category did not interact with age [$F(2, 436) = .14$, $p = .868$, $\eta^2 < .01$, $BF_{01} = 19.42$].

To examine the calibration of participants' post-task predictions of fluency for each category if they were to try again (predictions subtracted from performance), a one-sample t-test confirmed that, across topics, participants were well-calibrated [$t(225) = .98$, $p = .330$, $d = .07$, $BF_{01} = 8.40$]. A 2 (age: young, old) x 3 (category) mixed ANOVA revealed a main effect of category [$F(2, 448) = 4.18$, $p = .016$, $\eta^2 = .01$, $BF_{10} = 1.11$] but there were no significant pairwise differences in calibration for animals ($M = .79$, $SD = 27.13$), foods ($M = -3.19$, $SD = 20.40$), and occupations ($M = -1.82$, $SD = 15.14$), and Bayes Factor did not support this main effect we do not further interpret this result. Additionally, results did not reveal a main effect of age [$F(1, 224) < .01$, $p = .952$, $\eta^2 < .01$, $BF_{01} = 5.80$] such that younger adults were similarly post-task calibrated ($M = -1.18$, $SD = 12.90$) for each category as older adults ($M = -1.71$, $SD = 22.12$), and trial did not interact with age [$F(2, 448) = 1.83$, $p = .162$, $\eta^2 < .01$, $BF_{01} = 5.84$]. Thus, despite initial overconfidence, after having completed the task, participants were now calibrated in their metacognitive predictions of performance.

To examine participants' post-task predictions of fluency for each category for the average younger adult (aged 18-30), a 2 (age: young, old) x 3 (category) mixed ANOVA revealed a main effect of category [$F(2, 446) = 6.22$, $p = .002$, $\eta^2 = .03$, $BF_{10} = 7.65$] such that participants believed that the average younger adult would recall more foods ($M = 34.00$, $SD = 18.24$) than occupations ($M = 28.54$, $SD = 26.91$), [$p_{bonf} = .003$, $d = .22$] but not more foods than animals ($M = 32.49$, $SD = 22.40$), [$p_{bonf} = .625$, $d = .08$]; additionally, there were no differences in predictions for the average younger adult for animals and occupations [$p_{bonf} = .090$, $d = .15$]. However, results did not reveal a main effect of age [$F(1, 223) = .18$, $p = .675$, $\eta^2 < .01$, $BF_{01} = 5.67$] such that younger adults predicted that the average younger adult would retrieve a similar number of exemplars ($M = 31.19$, $SD = 16.26$) on each category as older adults predicted ($M = 32.21$, $SD = 20.66$), and trial did not interact with age [$F(2, 446) = 1.97$, $p = .140$, $\eta^2 = .01$, $BF_{01} = 5.24$].

To examine participants' post-task predictions of fluency for each category for the average older adult (aged 65-85), a 2 (age: young, old) x 3 (category) mixed ANOVA revealed a main effect of category [$F(2, 436) = 9.99$, $p < .001$, $\eta^2 = .04$, $BF_{10} > 100$] such that participants believed that the average older adult would recall more foods ($M = 38.91$, $SD = 52.22$) than occupations ($M = 29.50$, $SD = 22.53$), [$p_{bonf} = .006$, $d = .21$] and animals ($M =

28.42, SD = 15.37), [p_{bonf} = .004, d = .22]; however, there were no differences in predictions for the average older adult for animals and occupations [p_{bonf} > .999, d = .06]. Additionally, results did not reveal a main effect of age [$F(1, 218)$ = 1.55, p = .214, η^2 = .01, BF_{01} = 3.67] such that younger adults predicted that the average older adult would retrieve a similar number of exemplars (M = 30.45, SD = 20.91) on each category as older adults predicted (M = 34.22, SD = 29.67), and trial did not interact with age [$F(2, 436)$ = 1.00, p = .369, η^2 < .01, BF_{01} = 13.52].

Finally, regarding participants' post-task predictions of fluency for each age group, we were interested in whether participants predicted a similar level of performance for the average younger and older adult or expected that older adults would perform worse than younger adults. However, a 2 (age of participant: young, old) x 2 (prediction for young or old) mixed ANOVA did not reveal any significant effects [all ps > .162]. Thus, on their post-task predictions of performance for the average younger and older adult, participants did not expect any age-related differences.

Discussion

In Experiment 2, we tested both younger and older adults on three trials of a semantic fluency task to examine age-related differences in the calibration of their metacognitive predictions and if participants become better calibrated with increased task experience. Results did not reveal differences in fluency, originality, predictions of fluency, or calibration as a function of trial or age. However, older adults' postdictions indicated that they believed that they were less original than younger adults (the older the participant, the less original they rated their responses).

On the post-task questionnaire, there were no age-related differences in terms of participants' predictions of performance for each category if they were to try again and there were no age differences in predictions across categories. However, despite initial overconfidence, participants' predictions of their performance if they were to try again were calibrated with their actual performance. Additionally, there were no age-related differences in terms of participants' predictions of fluency performance for the average younger adult (aged 18-30) or the average older adult (aged 65-85). Finally, across participants, predictions of fluency and actual performance strongly positively correlated, indicating some metacognitive awareness of performance (although participants were generally underconfident) but there was not a significant relationship between originality scores and originality postdictions, indicating no metacognitive awareness of originality. Thus, although participants may have gained some metacognitive insight as to their ability to retrieve exemplars from semantic memory at the end of the task, younger and older adults are generally poor at monitoring their fluency and originality.

General Discussion

In the present study, we investigated age-related similarities and differences in the metacognitive awareness of fluency and originality. Younger and older adults engaged in several trials of a commonly used semantic fluency task and were asked to rate the originality of their responses. Using statistical infrequency as an index of originality, we

tracked the uniqueness of participants' responses and had participants judge their originality to see if they were aware of the qualitative aspects of their responses. In terms of fluency performance, we expected participants to be initially overconfident but to become better calibrated with increased task experience. We also expected older adults to be less metacognitively accurate in their originality judgments but for both younger and older adults to become more metacognitively accurate in their judgments of originality with increased task experience.

Results revealed that our younger adult sample was greatly underconfident in their performance, despite some metacognitive awareness of their ability to retrieve information from semantic memory (predictions and performance were correlated). Additionally, younger adults generally became better calibrated with task experience. There were no significant age-related differences in participants' metacognitive predictions and postdictions regarding originality, although older adults believed that they were less original than younger adults and became less original on later trials. These findings may fit with a broader literature that has found a lack of age differences in metacognition for verbal skills such as proofreading (Hargis et al., 2017) or proper name recall (Hargis et al., 2020) but shows that older adults may believe that their access to original verbal knowledge declines in older age. We note that these findings were obtained when participants had access to the output of their memory retrieval operations and different patterns may arise if participants make these judgments without being able to view their output.

The present study suggests that both younger and older adults display underconfidence in their retrieval fluency ability, revealing a "skilled and unaware" effect as well as a "memory-for-past-tests heuristic" whereby participants were underconfident at first but became better calibrated after experiencing the task. Little work has found underconfident high-performers (see Dunning, 2011) and this underconfidence fits with some prior work that has found similar effects in younger and older adults using episodic learning tasks (Tauber & Rhodes, 2012). Additionally, the lack of age differences may have implications for delineating healthy from pathological aging, since metacognition is often impaired in patients with Alzheimer's Disease or other neurological conditions (see Eslinger et al., 2005; McGlynn & Kaszniak, 1991). Moreover, rather than tasks that assess episodic memory, the current tasks measure verbal fluency, supporting work suggesting that vocabulary and other forms of semantic memory are often fairly intact in older age (see Ben-David & Erel, 2015; Kavé & Halamish, 2015). Furthermore, older adults may have reduced inhibition that benefits performance in these types of verbal free-response tasks (cf. Healey et al., 2008; Kim et al., 2007), and in the present task, reduced inhibition may enhance semantic access in terms of outputting more exemplars in older age.

Participants may initially poorly judge their originality due to the influence of recency effects such that the last few items recalled are more likely to be original items (cf., Murphy & Castel, 2021; Sidi et al., 2020). Specifically, both age groups may initially use the last few emitted items as the basis for their originality judgments, leading to a poor evaluation of their overall originality. However, on later trials, younger adults may monitor their originality throughout the retrieval process, leading to a better awareness of the originality of their responses (see Koriat, 1993 for similar work asking participants to provide feeling of

knowing judgments). Thus, having some experience with the task, and greater awareness of the retrieval dynamics from semantic memory, can inform metacognition (although originality scores and ratings did not correlate in Experiment 2).

In our sample of younger and older adults (from Amazon's Mechanical Turk), we did not observe the age-related differences in fluency and originality observed in some prior work (e.g., Murphy & Castel, 2021; Rodríguez-Aranda & Martinussen, 2006; Troyer et al., 1997), although some of this same prior work has shown age-related similarities for "younger" older adults, with a mean age of 62.44 years (Murphy & Castel, 2021). We acknowledge that the older sample tested in the present work may represent a relatively "younger" older adult sample (68.31 years old), who may also not have experienced the stress that can accompany in-person laboratory testing of memory (Barber & Mather, 2014) since they were participating remotely (similar to our prior relevant work, see Murphy & Castel, 2021). Thus, these participants and conditions may not generalize to all older adults but do provide useful insights as the sample is comparable in age to prior work and may illustrate how healthy older adults function in more naturalistic settings. In Alzheimer's disease or other forms of dementia, metacognition for verbal fluency and originality may be impaired, and this would provide an important future direction in terms of delineating healthy and pathological aging.

The present work provides a novel investigation of metacognition in verbal fluency for younger and relatively healthy older adults. Recent work has examined some of these effects in younger adults with an idea generation task (Sidi et al., 2020), and the present work extends this to verbal fluency and shows that younger and older adults show comparable effects in terms of performance and how their metacognitive judgments are underconfident. We also found that older adults are less confident in assessing the originality of their responses. However, this may also arise due to their frequent use of verbal knowledge and not feeling that access to some information is novel as a result of a lifetime of experience with this information, although future research is needed to more fully explore the precise basis for these judgments. Additionally, since originality postdictions were on a Likert scale, we were unable to examine over- or under-confidence in the originality of one's responses; future work may benefit from asking participants to estimate the number of original exemplars they retrieved.

In terms of participants' underconfidence, it may be that in general, they did not want to set lofty expectations and/or they did not know how much accessing semantic memory would yield. Previous work using episodic memory tasks has demonstrated some overconfidence, particularly in older adults (e.g., McGillivray & Castel, 2011). Based on these findings from episodic memory tasks, we expected some initial levels of overconfidence. However, if semantic retrieval monitoring follows different mechanisms than those used in episodic memory, then overconfidence may not be present. As such, since participants were largely underconfident, monitoring retrieval from semantic memory may follow different mechanisms/have different properties than episodic memory and, when retrieving items from semantic memory, we know more than we know.

Moreover, when one fails to retrieve a familiar item, this can arise from tip-of-the-tongue states (Brod et al., 2013; Brown, 1991; Brown & McNeill, 1966; Burke et al., 1991; Schwartz & Metcalfe, 2011). Participants may also experience a "feeling of knowing" whereby items are available in semantic memory but are currently inaccessible (Hart, 1965; Koriat, 1993, 1995; see Schwartz, 1999; Schwartz & Metcalfe, 2014, for reviews). In the present task, these states may have contributed to participants' underestimation of how much information they could retrieve from semantic memory such that participants think that exemplars are less accessible than they are, or they underestimate their ability to overcome tip-of-the-tongue states to retrieve items.

The present work revealed some interesting age-related similarities and differences in terms of originality and judgments regarding originality. While there were no age differences in originality, older adults' postdictions indicated that they believed that they were less original than younger adults. These results may suggest that despite younger and older adults performing similarly in terms of originality, one's awareness of originality may be reduced with age. While the relationship between age and creativity is much debated (Carpenter et al., 2020; Simonton, 1990), it may be that older adults perceive less originality due to greater experience, verbal knowledge and/or vocabulary, and comparisons to a wider set of experiences. In contrast, younger adults may fall into the "better than average" illusion (Dunning et al., 2004; Hargis et al., 2020) in terms of reporting on originality, although further research is needed to investigate this in a more systematic manner. In the present task, since we were interested in participants' global assessment of originality, we asked participants to provide a general assessment of their originality performance rather than a prediction of the number of original responses. However, this global or holistic measure differs from an assessment that measures the number of original responses and future research could compare these types of metacognitive operations (judgments of overall originality versus the quantity of original responses) as they may rely on different mechanisms.

We observed a few differences between Experiments 1 and 2. For example, in Experiment 2, there was no significant improvement in calibration as a function of task experience but there was in Experiment 1. Additionally, fluency scores and the originality judgments for the younger adults appear to differ between the two experiments and these differences may be attributable to the different samples used in each experiment. Specifically, in Experiment 1 participants were undergraduate students from a large public university whereas in Experiment 2 participants were recruited online. Undergraduate students may have better access to semantic memory since they are amid their educational careers (i.e., undergraduate students are frequently completing examinations that require them to retrieve information from semantic memory) whereas participants from other sources may retrieve information from semantic memory less frequently (and younger adults in Experiment 1 recalled around 20 more exemplars per trial, on average). Future work should examine these effects in more diverse samples with different educational backgrounds.

It is important to note that the present study employed a paradigm in which participants' responses remained on-screen during the entirety of the retrieval period while some other work uses methods that do not allow participants to see their responses. This procedural

difference is important in that it provides access to retrieval output, but since one of our main research objectives was to examine how younger and older adults assess the originality of information that is retrieved from semantic memory, we wanted to ensure that participants could view their responses during retrieval so that this measure was not biased/influenced by retrieval failure which may be more pronounced in older adults. Again, we elected to use this type of procedure so that younger and older adults would be more aware of their responses and subsequently be better able to evaluate the originality of their responses. Specifically, if younger and older adults did not have access to their responses, we could not rule out explanations based on poorer access to the retrieved material which could potentially drive age-related differences in assessments of originality. However, this modified procedure may change the way participants monitor previously retrieved items to avoid repetitions and search for items in semantic memory—factors that may influence retrieval processes. Performance impairments in older adults sometimes arise from repetition errors or a lack of switching (generating cues for a new category of exemplars; Troyer et al., 1997), thus, the lack of age-related differences in retrieval fluency and metacognition observed here may be due to this procedural difference. Future work may benefit from examining metacognitive predictions of performance when participants cannot see previously retrieved items and also by assessing originality judgments when participants must rely on memory for their responses to determine if there may be age-related differences in different forms of global originality assessments.

In sum, while we found that predictions of semantic fluency and actual performance were generally strongly positively correlated, indicating some metacognitive awareness of performance in both younger and older adults, both groups underestimated their performance and sometimes failed to accurately assess the originality of their responses. This suggests that both younger and older adults may gain access to more than one is aware, which can help in better understanding age-related changes in the metacognition of creativity, learning, and communication.

Acknowledgments

References

Baker JMC, Dunlosky J, & Hertzog C (2010). How accurately can older adults evaluate the quality of their text recall? The effect of providing standards on judgment accuracy. Applied Cognitive Psychology, 24, 134–147.

Barber SJ, & Mather M (2014). Stereotype threat in older adults: When and why does it occur and who is most affected? In Verhaeghen P & Hertzog C (Eds.), The Oxford handbook of emotion, social cognition, and problem solving in adulthood (pp. 302–319). Oxford University Press.

Ben-David BM, & Erel H (2015). "Older is always better": Age-related differences in vocabulary scores across 16 years. Psychology and Aging, 30, 856–862. [PubMed: 26652725]

Benedek M, Fink A, & Neubauer AC (2006). Enhancement of ideational fluency by means of computer-based training. Creativity Research Journal, 18, 317–328.

Boot N, Baas M, Van Gaal S, Cools R, & De Dreu CKW (2017). Creative cognition and dopaminergic modulation of fronto-striatal networks: Integrative review and research agenda. Neuroscience & Biobehavioral Reviews, 78, 13–23. [PubMed: 28419830]

Brod G, Werkle-Bergner M, & Shing YL (2013). The influence of prior knowledge on memory: A developmental cognitive neuroscience perspective. Frontiers in Behavioral Neuroscience, 7, 139. [PubMed: 24115923]

Brown AS (1991). A review of the tip-of-the-tongue experience. Psychological Bulletin, 109, 204–223. [PubMed: 2034750]

Brown R, & McNeill D (1966). The "tip of the tongue" phenomenon. Journal of Verbal Learning & Verbal Behavior, 5, 325–337.

Burke DM, MacKay DG, Worthley JS, & Wade E (1991). On the tip of the tongue: What causes word finding failures in young and older adults? Journal of Memory and Language, 30, 542–579.

Carpenter SM, Chae RL, & Yoon C (2020). Creativity and aging: Positive consequences of distraction. Psychology and Aging, 35, 654–662. [PubMed: 32744848]

Coane JH, & Umanath S (2019). I don't remember vs. I don't know: Phenomenological states associated with retrieval failures. Journal of Memory and Language, 107, 152–168.

Dodson CS, Bawa S, & Krueger LE (2007). Aging, metamemory, and high-confidence errors: A misrecollection account. Psychology and Aging, 22, 122–133. [PubMed: 17385989]

Dumas D, & Dunbar K (2014). Understanding fluency and originality: A latent variable perspective. Thinking Skills and Creativity, 14, 56–67.

Dunlosky J, & Metcalfe J (2008). Metacognition: A textbook for cognitive, educational, lifespan and applied psychology. Thousand Oaks, CA: SAGE.

Dunlosky J, Mueller ML, & Thiede KW (2016). Methodology for investigating human metamemory: Problems and pitfalls. In Dunlosky J & Tauber SK (Eds.), Oxford library of psychology. The Oxford handbook of metamemory (p. 23–37). Oxford University Press.

Dunning D (2011). The Dunning-Kruger effect: On being ignorant of one's own ignorance. In Advances in experimental social psychology (Vol. 44, pp. 247–296). San Diego, CA: Elsevier.

Dunning D, Heath C, & Suls JM (2004). Flawed self-assessment: Implications for health, education, and the workplace. Psychological Science in the Public Interest, 5, 69–106. [PubMed: 26158995]

Eslinger PJ, Dennis K, Moore P, Hauck R, & Grossman M (2005). Metacognitive deficits in frontotemporal dementia. Journal of Neurology, Neurosurgery & Psychiatry, 76, 1630–1635. [PubMed: 16291884]

Gruenewald PJ, & Lockhead GR (1980). The free recall of category examples. Journal of Experimental Psychology: Human Learning and Memory, 6, 225–240.

Halamish V, McGillivray S, & Castel AD (2011). Monitoring one's own forgetting in younger and older adults. Psychology and Aging, 26, 631–635. [PubMed: 21463057]

Hargis MB, Whatley MC, & Castel AD (2020). Remembering proper names as a potential exception to the better-than-average effect in younger and older adults. Psychology and Aging, 35, 497–507. [PubMed: 32352805]

Hargis MB, Yue CL, Kerr T, Ikeda K, Murayama K, & Castel AD (2017). Metacognition and proofreading: the roles of aging, motivation, and interest. Aging, Neuropsychology, and Cognition, 24, 216–226.

Hart JT (1965). Memory and the feeling-of-knowing experience. Journal of Educational Psychology, 56, 208–216. [PubMed: 5825050]

Healey MK, Campbell KL, & Hasher L (2008). Cognitive aging and increased distractibility: Costs and potential benefits. Progress in Brain Research, 169, 353–363. [PubMed: 18394486]

Henry JD, & Crawford JR (2004). A meta-analytic review of verbal fluency performance following focal cortical lesions. Neuropsychology, 18, 284–295. [PubMed: 15099151]

Hertzog C, & Dunlosky J (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. Current Directions in Psychological Science, 20, 167–173. [PubMed: 24478539]

Hertzog C, Kidder DP, Powell-Moman A, & Dunlosky J (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? Psychology and Aging, 17, 209–225. [PubMed: 12061407]

Hertzog C, Sinclair SM, & Dunlosky J (2010). Age differences in the monitoring of learning: Cross-sectional evidence of spared resolution across the adult life span. Developmental Psychology, 46, 939–948. [PubMed: 20604613]

Hess TM (2005). Memory and aging in context. Psychological Bulletin, 131, 383–406. [PubMed: 15869334]

Hocevar D (1979). A comparison of statistical infrequency and subjective judgment as criteria in the measurement of originality. Journal of Personality Assessment, 43, 297–99. [PubMed: 16367009]

Hocevar D (1980). Intelligence, divergent thinking, and creativity. Intelligence, 4, 25–40.

Kass RE, & Raftery AE (1995). Bayes factors. Journal of the American Statistical Association, 90, 773–795.

Kavé G, & Halamish V (2015). Doubly blessed: Older adults know more vocabulary and know better what they know. Psychology and Aging, 30, 68–73. [PubMed: 25602490]

Kim S, Hasher L, & Zacks RT (2007). Aging and a benefit of distractibility. Psychonomic Bulletin & Review, 14, 301–305.

Koriat A (1993). How do we know that we know? The accessibility model of the feeling of knowing. Psychological Review, 100, 609–639. [PubMed: 8255951]

Koriat A (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. Journal of Experimental Psychology: General, 124, 311–333.

Koriat A (2016). Metacognition: Decision-making processes in self-monitoring and self regulation. In Keren G, & Wu G (Eds.), The Wiley Blackwell handbook of judgment and decision making, Volume 1 (pp. 356–379). Malden, MA: Wiley- Blackwell.

McGillivray S, & Castel AD (2011). Betting on memory leads to metacognitive improvement by younger and older adults. Psychology and Aging, 26, 137–142. [PubMed: 21417541]

McGlynn SM, & Kaszniak AW (1991). When metacognition fails: Impaired awareness of deficit in Alzheimer's disease. Journal of Cognitive Neuroscience, 3, 183–187. [PubMed: 23972092]

Milgram RM, & Milgram NA (1976). Creative thinking and creative performance in Israeli students. Journal of Educational Psychology, 3, 255–259.

Moore DA, & Healy PJ (2008). The trouble with overconfidence. Psychological Review, 115, 502–517. [PubMed: 18426301]

Murphy DH, & Castel AD (2021). Age-related similarities and differences in the components of semantic fluency: Analyzing the originality and organization of retrieval from long-term memory. Aging, Neuropsychology and Cognition, 28, 748–761.

Murphy DH, & Castel AD (2022). Differential effects of proactive and retroactive interference in value-directed remembering for younger and older adults. Psychology and Aging.

Murphy DH, & Castel AD (2022). The role of attention and ageing in the retrieval dynamics of value-directed remembering. Quarterly Journal of Experimental Psychology, 75, 954–968.

Nelson TO (1996). Consciousness and metacognition. American Psychologist, 51, 102–116.

Nelson TO, & Narens L (1990). Metamemory: A theoretical framework and new findings. Psychology of Learning and Motivation, 26, 125–141.

Park DC, & Festini SB (2017). Theories of memory and aging: A look at the past and a glimpse of the future. Journals of Gerontology: Psychological Sciences, 72, 82–90.

Rhodes MG (2016). Judgments of learning. In Dunlosky J and Tauber SK (Eds.), The Oxford handbook of metamemory (pp. 65–80). New York: Oxford University Press.

Rhodes MG, & Castel AD (2008). Metacognition and part-set cuing: Can interference be predicted at retrieval? Memory & Cognition, 36, 1429–1438.

Rodríguez-Aranda C, & Martinussen M (2006). Age-related differences in performance of phonemic verbal fluency measured by controlled oral word association task (COWAT): A meta-analytic study. Developmental Neuropsychology, 30, 697–717. [PubMed: 16995832]

Rosen VM, & Engle RW (1997). The role of working memory capacity in retrieval. Journal of Experimental Psychology: General, 126, 211–227. [PubMed: 9281831]

Runco MA, & Charles R (1993). Judgments of originality and appropriateness as predictors of creativity. Personality and Individual Differences, 15, 537–546.

Salthouse TA (2010). Selective review of cognitive aging. Journal of the International Neuropsychological Society, 16, 754–760. [PubMed: 20673381]

Schwartz BL (1999). Sparkling at the end of the tongue: The etiology of tip-of-the-tongue phenomenology. Psychonomic Bulletin & Review, 6, 379–393. [PubMed: 12198776]

Schwartz BL, & Metcalfe J (2011). Tip-of-the-tongue (TOT) states: Retrieval, behavior, and experience. Memory & Cognition, 39, 737–749. [PubMed: 21264637]

Schwartz BL, & Metcalfe J (2014). Tip-of-the-tongue (TOT) states: Mechanisms and metacognitive control. In Schwartz BL, & Brown AS (Eds.). Tip-of-the-Tongue states and related phenomena (pp. 15–31). Cambridge, UK: Cambridge University Press.

Sidi Y, Torgovitsky I, Soibelman D, Miron-Spektor E, & Ackerman R (2020). You may be more original than you think: Predictable biases in self-assessment of originality. Acta Psychologica, 203, 103002. [PubMed: 32004640]

Silvia PJ (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). Creativity Research Journal, 20, 34–39.

Silvia PJ (2011). Subjective scoring of divergent thinking: Examining the reliability of unusual uses, instances, and consequences tasks. Thinking Skills and Creativity, 6, 24–30.

Simonton DK (1990). Creativity and wisdom in aging. Handbook of the Psychology of Aging, 3, 320–329.

Souchay C, Moulin CJA, Clarys D, Taconnat L, & Isingrini M (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. Consciousness and Cognition: An International Journal, 16, 769–784.

Sowden PT, Pringle A, & Gabora L (2015). The shifting sands of creative thinking: Connections to dual-process theory. Thinking & Reasoning, 21, 40–60.

Stine-Morrow EAL, Shake MC, Miles JR, & Noh SR (2006). Adult age differences in the effects of goals on self-regulated sentence processing. Psychology and Aging, 21, 790–803. [PubMed: 17201498]

Tauber SK, & Rhodes MG (2012). Multiple bases for young and older adults' judgments of learning in multitrial learning. Psychology and Aging, 27, 474–483. [PubMed: 21942899]

Troster AI, Fields JA, Testa JA, Paul RH, Blanco CR, Hames KA, Salmon DP, & Beatty WW (1998). Cortical and subcortical influences on clustering and switching in the performance on verbal fluency tasks. Neuropsychologia, 36, 295–304. [PubMed: 9665640]

Troyer AK, Moscovitch M, & Winocur G (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. Neuropsychology, 11, 138–146. [PubMed: 9055277]

Unsworth N, Spillers GJ, & Brewer GA (2011). Variation in verbal fluency: A latent variable analysis of switching, clustering, and overall performance. Quarterly Journal of Experimental Psychology, 64, 447–466.

Witherby AE, Tauber SK, Rhodes MG, & Castel AD (2019). Aging and forgetting: Forgotten information is perceived as less important than is remembered information. Psychology and Aging, 34, 228–241. [PubMed: 30550310]
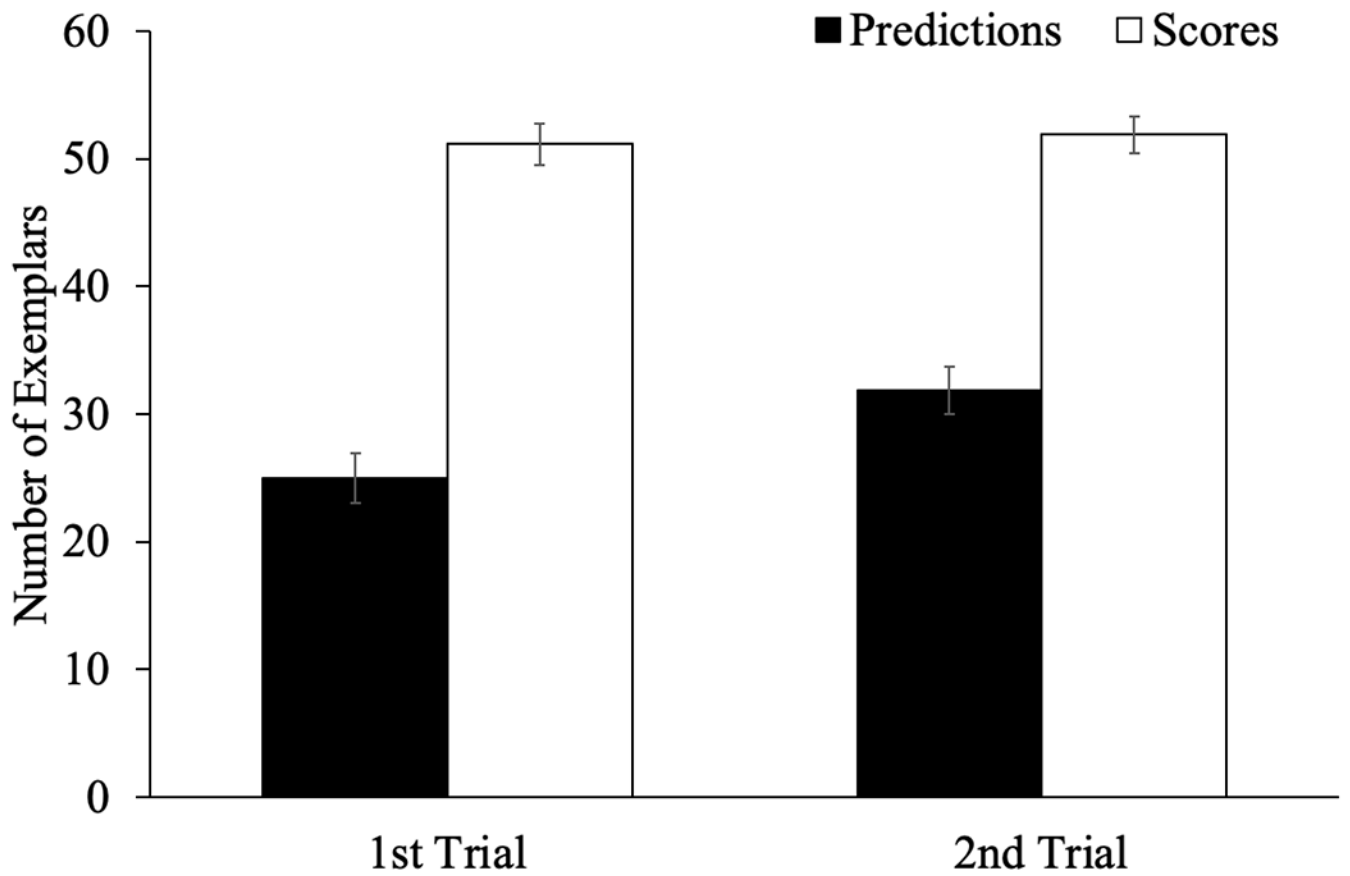
Figure 1.
Fluency predictions and scores as a function of task experience for younger adults in Experiment 1. Error bars reflect the standard error of the mean.
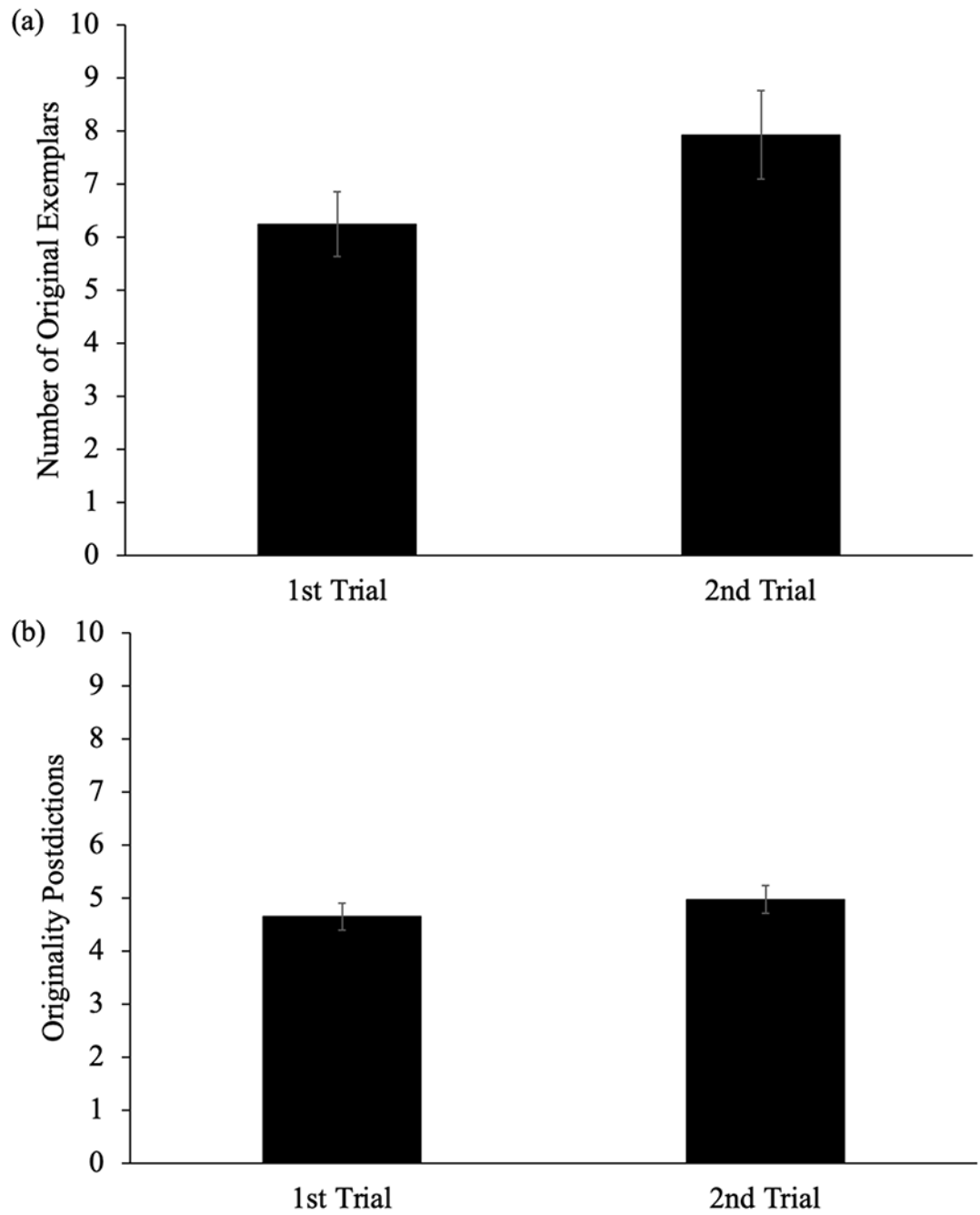
Figure 2.
Number of original exemplars retrieved (a) and originality postdictions (b) as a function of task experience for younger adults in Experiment 1. Error bars reflect the standard error of the mean. Note that originality scores and postdictions are on different scales and should not be compared.

Figure 3.

Fluency predictions and scores as a function of task experience for younger adults (a) and older adults (b) in Experiment 2. Error bars reflect the standard error of the mean.

Figure 4.
Number of original exemplars retrieved (a) and originality postdictions (b) as a function of task experience and age in Experiment 2. Error bars reflect the standard error of the mean. Note that originality scores and postdictions are on different scales and should not be compared.
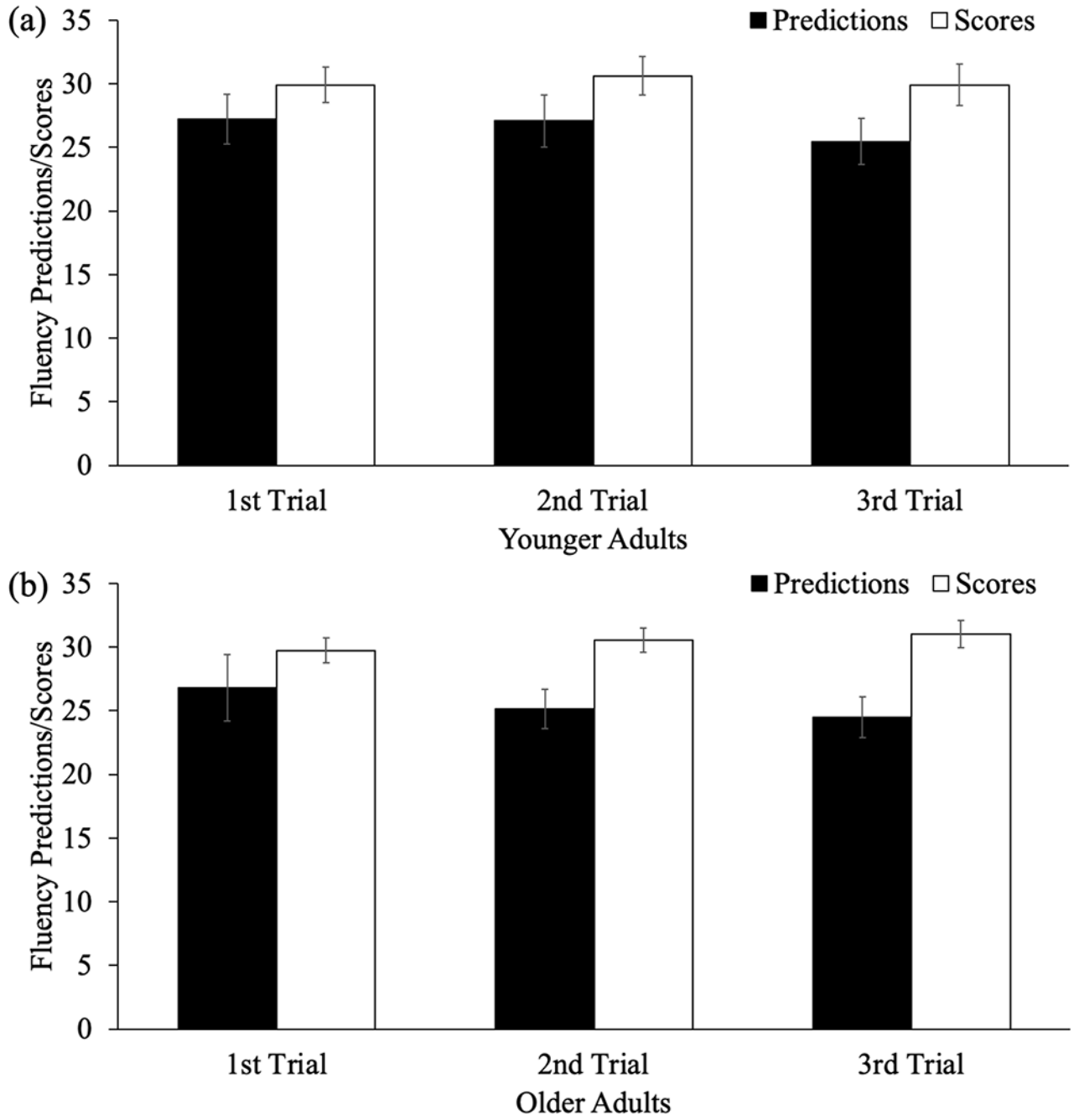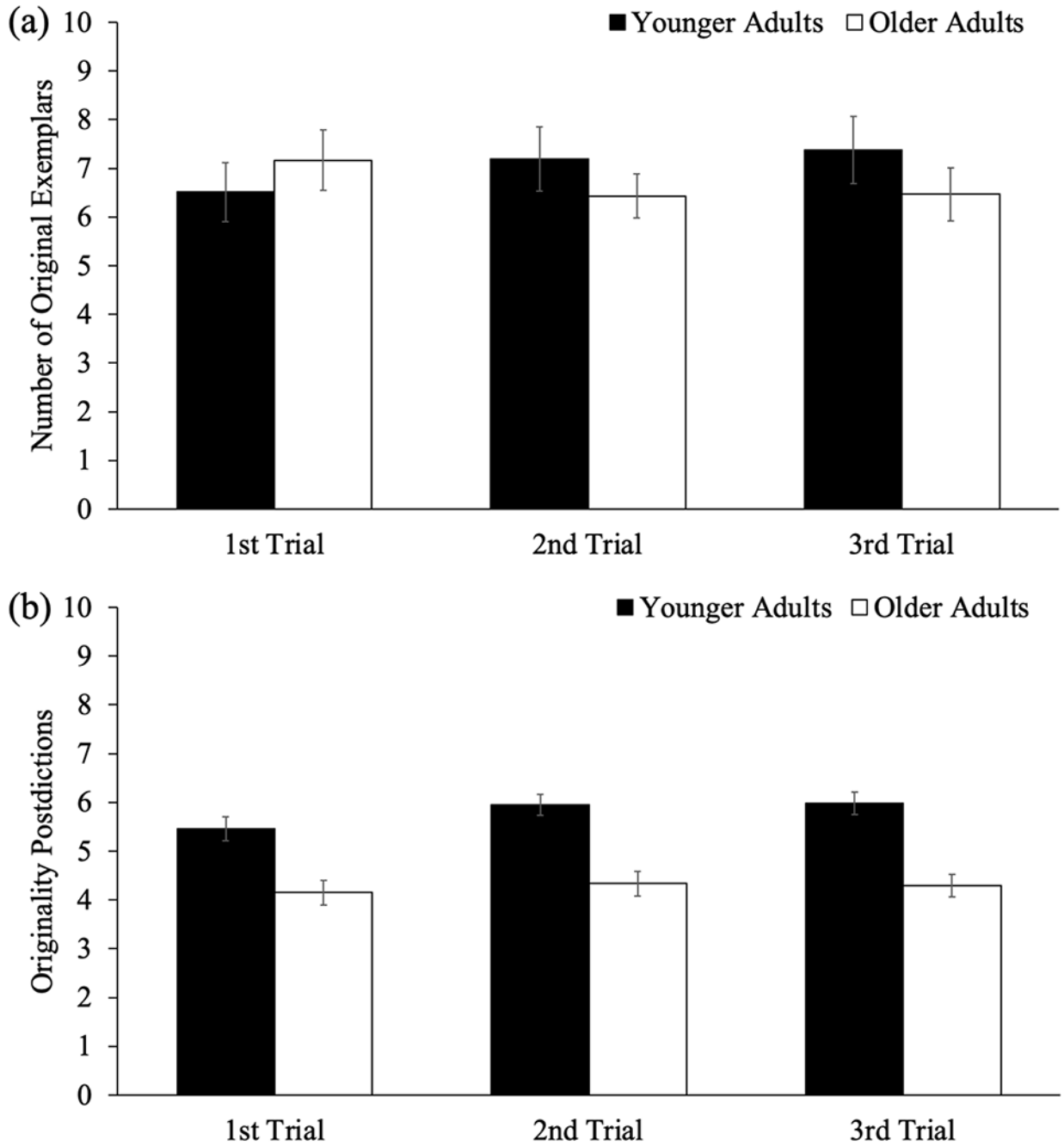
Table 1.

The most frequently generated items and examples of items outputted only once for each category in the current study.

Common Animals	Original Animals	Common Foods	Original Foods	Common Occupations	Original Occupations
Dog	Axolotl	Bread	Broccolini	Doctor	Stunt double
Cat	Adder	Apples	Challah	Nurse	Taste tester
Lion	Caracal	Milk	Clams	Teacher	Ornithologist
Tiger	Dhole	Cheese	Daikon	Lawyer	CIA agent
Pig	Echidna	Bananas	Grapefruit	Police officer	Animal trainer
Giraffe	Jaguarundi	Chicken	Mint	Waiter	Music teacher
Horse	Mola	Ice cream	Mussels	Actor	Nun
Cow	Monkfish	Tomatoes	Pierogi	Plumber	Geneticist
Elephant	Nautilus	Eggs	Prosciutto	Firefighter	Ventrilloquist
Whale	Okapi	Chips	Sorbet	Accountant	Transcriptionist

Table 2.

Pearson correlations between fluency and originality performance and metacognitive predictions/postdictions on each trial in Experiment 1.

Measure	1	2	3	4	5	6	7	8
1) 1st Trial Fluency Prediction	—							
2) 1st Trial Fluency	.370 ***	—						
3) 1st Trial Originality	.211 *	.733 ***	—					
4) 1st Trial Originality Postdiction	.064	.138	.108	—				
5) 2nd Trial Fluency Prediction	.588 ***	.399 ***	.203 *	.170	—			
6) 2nd Trial Fluency	.312 **	.637 ***	.432 ***	.219 *	.399 ***	—		
7) 2nd Trial Originality	.241 *	.407 ***	.401 ***	.233 *	.288 ***	.697 ***	—	
8) 2nd Trial Originality Postdiction	.100	−.018	.036	.616 ***	.041	.188	.194 *	—

Note:

*
= $p < .05$,

**
= $p < .01$,

***
= $p < .001$.

Table 3.

Pearson correlations between fluency and originality performance and metacognitive predictions/postdictions on each trial in Experiment 2.

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1) 1st Trial Fluency Prediction	—												
2) 1st Trial Fluency	.277***	—											
3) 1st Trial Originality	.234***	.659***	—										
4) 1st Trial Originality Postdiction	−.063	−.106	−.022	—									
5) 2nd Trial Fluency Prediction	.623***	.338***	.409***	−.041	—								
6) 2nd Trial Fluency	.185**	.738***	.662***	−.165*	.438***	—							
7) 2nd Trial Originality	.226***	.633***	.456***	−.095	.346***	.672***	—						
8) 2nd Trial Originality Postdiction	−.100	−.163*	−.053	.688***	−.086	−.181**	−.010	—					
9) 3rd Trial Fluency Prediction	.603***	.393***	.382***	−.098	.590***	.311***	.384***	−.138*	—				
10) 3rd Trial Fluency	.201**	.748***	.656***	−.169*	.341***	.755***	.601***	−.207**	.468***	—			
11) 3rd Trial Originality	.187**	.607***	.457***	−.146*	.285***	.611***	.579***	−.161*	.338***	.701***	—		
12) 3rd Trial Originality Postdiction	−.131*	−.104	.029	.665***	−.049	−.095	.063	.754***	−.097	−.141*	−.050	—	
13) Age	−.015	−.008	.046	−.241***	−.065	−.011	−.069	−.304***	−.032	.027	−.082	−.326***	—

Note:

*
= $p < .05$,

**
= $p < .01$,

***
= $p < .001$.