



Monitoring and control processes in mock witnesses in under-represented non-WEIRD samples with high or low educational level

Karlos Luna¹ · Sara Cadavid² · Inés Botía²

Accepted: 12 March 2022
© The Psychonomic Society, Inc. 2022

Abstract

A popular model proposes that metamemory is based on two processes, monitoring and control. The first examines memories and evaluates their quality and the second uses that information to decide on the most appropriate course of action. Monitoring and control processes have been studied mostly with university students, which raises the question of how well do they work in groups of people from under-represented samples such as people with a low educational level. In this research, we tested the monitoring and control processes of three groups of participants from a non-WEIRD (Western, Educated, Industrialized, Rich and Democratic) country (Colombia). Two groups of adults (aged 30–55 years) living in urban or rural areas and with a low educational level and a group of Colombian university students watched a bank robbery video and answered cued recall questions. To measure monitoring ability, participants rated their confidence that they had produced the correct answer, and to measure control they indicated whether they preferred to report or withhold the response were they in a trial. Results showed that the three groups had a functional ability to monitor their memories and control their behaviour, and that university students had better memory and metamemory than the two low education groups. The results support the concept that the basic metamemory processes of monitoring and control are functional in different groups of individuals, but the differences between groups highlight the need to test the generalizability of cognitive processes and phenomena across individuals.

Keywords Metamemory · Monitoring · Control · Non-WEIRD · Educational level · Eyewitness memory

Introduction

Metamemory refers to the knowledge we have about our memory functioning and to what we do with that knowledge. A popular and influential metamemory model proposes two basic processes, monitoring and control (Nelson & Narens, 1990, 1994). Monitoring refers to the ability to examine our memories, for example, to check whether they are correct or incorrect. Control refers to the behavioural changes that result from the information obtained by the monitoring

process, for example, the decision on whether to provide an answer or opt for a "don't know" response.

Besides its theoretical interest, the capacity to monitor memory and control behaviour also has a considerable applied interest. For example, in forensic settings it is important that witnesses monitor their memories and rate correct and incorrect responses with different confidence levels (e.g., Loftus et al., 1989; Luna & Martín-Luengo, 2012). It is also important that witnesses' behaviour reflects their ability to monitor their memories, for example by reporting information with high chances of being correct, that is, rated with high confidence, and withholding information with low chances of being correct, that is, rated with low confidence (e.g., Evans & Fisher, 2011). Similarly, basic monitoring processes are relevant in educational settings, in which students have to monitor their learning process and make decisions on the best learning strategy and how to allocate their study time (for a review, see Soderstrom et al., 2016). Also, metamemory and the monitoring-control model have

✉ Karlos Luna
klunao@unal.edu.co

¹ Department of Psychology, Universidad Nacional de Colombia, Bogotá, Colombia

² School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia

proven useful to study mental disorders such as schizophrenia (Moritz & Woodward, 2006; Moritz et al., 2006), autism, attention-deficit hyperactivity disorder (ADHD), depression, or obsessive-compulsive disorder (for a review, see Izaute & Bacon, 2016). The monitoring-control model has also been applied to areas traditionally not close to psychology, such as cybersecurity (Luna, 2019).

In line with its theoretical and applied relevance, the monitoring-control model has received substantial empirical support (see, e.g., Dunlosky & Tauber, 2016). However, most of it comes from a particular group of people: university students from WEIRD (Western, Educated, Industrialized, Rich and Democratic) countries. This is problematic because WEIRD samples are unusual in many psychological and behavioural dimensions, and are, thus, not representative of the human species (Henrich et al., 2010). In addition, most research in cognitive science is conducted with university students, which are a more homogeneous group than the general population (Peterson, 2001). Thus, reliance on WEIRD samples of mostly university students limits the generalizability of the conclusions obtained in cognitive sciences (Henrich et al., 2010; Rad et al., 2018; Tiokhin et al., 2019; for a review, see the special issue in *Evolution and Human Behavior* edited by Apicella et al., 2020).

Related to the generalizability issue and focusing now on metamemory, research has identified variables and situations in which metamemory does not work as expected or is not functional in the sense of not helping people complete their tasks successfully (e.g., Luna & Martín-Luengo, 2014; Peng & Tullis, 2021; Rhodes & Castel, 2008; any situation that would not fit into the "pristine conditions" of eyewitness identification, see Wixted & Wells, 2017). These arguments raise the question of how well the monitoring and control processes work in groups of people with different characteristics compared with the widely studied university student populations from WEIRD countries. In this research, we examined the functioning of basic metamemory processes in groups of people from under-represented samples from a non-WEIRD country.

One relevant cognitive characteristic of university students is their educational level. Educational level is known to affect different cognitive functions in a healthy population, for example, verbal memory (Argento et al., 2015), visual memory (Rosselli & Ardila, 2003), working memory (Zarantonello et al., 2020), or sensory tasks (Stratta et al., 2001). Along these lines, Murre et al. (2013) found in a sample of 28,000 Dutch participants that people with only primary studies performed worse in both verbal and visual memory tasks than people with secondary or higher education. Consistently, a higher educational level has been associated, in older adults, with self-reports of having better metamemory, measured with the Metamemory in Adulthood Inventory (Guerrero-Sastoque et al., 2021). However,

a study on memory for odours found that only older adults with graduate studies have better metamemory than older adults with bachelor or high school degrees, with no differences between the latter two (Szajer & Murphy, 2013). Then, if educational level is linked to better metamemory, it may be so only for people with the highest educational degrees. In that case, the observed metamemory improvement may not be an effect of higher educational level but of individual differences that make some people enter graduate school.

In contrast, other studies have found that educational level is not related to metamemory. For example, Quattropani et al. (2016) found no differences between educational levels in healthy adults with the Metacognitions Questionnaire 30 (MCQ-30), a self-report questionnaire that measures metacognitive beliefs and processes. Similarly, Soler and Ruiz (1996) found that educational level did not affect the use of mnemonic techniques such as mental rehearsal, but that it did affect the use of other strategies such as short-term repetition. However, participants in that study were secondary students aged 15 or 16 years and university students aged 21 years, and thus educational level and age could be confounded. Therefore, the results from that study should be interpreted with caution because developmental issues may have been at play.

In sum, several research lines show an apparent effect of educational level on different cognitive processes. However, the limited amount of research on the effect of educational level on metamemory shows mixed results. Thus, the question of whether educational level affects metamemory remains unsolved. To answer this question, we tested the monitoring and control abilities of adults with different educational levels. Specifically, our participants were two groups of adults with low educational levels living in urban or rural areas and a control group of university students, included for comparison purposes. To our knowledge, this is the first research in which people with different educational levels (in either WEIRD or non-WEIRD countries) participated in an experiment about metamemory related to specific memories (and not general beliefs about memory functioning or the use of mnemonic strategies, as in metamemory questionnaires). Since the literature did not show a clear effect of the educational level on metamemory, we tentatively expected no effect of educational level in monitoring and control tasks.

Monitoring and control processes have been studied at both encoding (through judgements of learning; e.g., Little & McDaniel, 2015; Luna et al., 2019) and retrieval (through confidence ratings; e.g., Arnold et al., 2013; Luna et al., 2011). We chose confidence ratings for two reasons: their suitability for our samples and their relevance to eyewitness memory. First, research with judgements of learning usually involves learning a list of words and then recalling it, but the use of verbal materials like those used in education may

provide university students with an advantage because of their higher experience with verbal materials. Also, people who are not used to study verbal materials may not be motivated to enrol in an experiment with such materials. Thus, we relied on a video as the to-be-remembered material. Typically, metamemory for video contents is studied with confidence ratings, so we used that measure in this research. Second, confidence ratings are relevant in eyewitness memory, an area in which there is debate over whether and under which conditions the monitoring and control processes work. For example, for years it was thought that the relationship between confidence and accuracy in eyewitness memory was weak for both event memory (e.g., Perfect et al., 1993, 2000) and identification studies (Brewer et al., 2002; Sporer et al., 1995). However, later research showed that metamemory was reliable even with eyewitness memory materials (e.g., Luna & Martín-Luengo, 2012). In identification studies, there is also a debate over the conditions that promote a strong or weak confidence-accuracy relationship (see Sauer et al., 2019). Thus, the effectiveness of metamemory processes should not be taken for granted, and eyewitness memory materials and confidence ratings provide a good opportunity to test that effectiveness.

In the experiment reported below, participants from a non-WEIRD country watched a bank robbery video and answered cued recall questions. Participants indicated their confidence on having provided the correct answer and whether they would like to report that particular answer were they witnesses in a trial. We expected that the three groups of participants would show functional monitoring and control. In other words, we expected that participants would be able to distinguish between correct and incorrect responses (i.e., monitoring) and that they would use that information to guide their decisions (i.e., control). In addition, we also expected that educational level would not affect the effectiveness of the basic metamemory processes.

Method

Participants and design

This research was approved by the local ethics committee. Our design was a 3 (group: university students with high educational level, urban participants with low educational level, rural participants with low educational level) manipulated between participants. We included two groups with low educational level living in different areas to add more variability to our sample. We made no predictions over the effect of place of living in metamemory. Luna and Martín-Luengo (2012) found that the difference between confidence for correct and incorrect responses (i.e., the simplest monitoring measure) was very large with eyewitness memory materials,

$d_{av}=2.51$. Thus, we relied on a similar sample to that used by Luna and Martín-Luengo (they had one single group of 53 participants). A total of 165 participants (104 females, mean age 33.32 years, $SD=10.78$) completed the experiment voluntarily.

There were 55 Colombian participants in each of the three groups and we set specific requirements for participation. University students were between 18 and 25 years old and were at least in their fourth semester of higher education (most undergraduate degrees in Colombia span ten semesters). We avoided the youngest students for two reasons: (1) to maximize the effect of education when compared with the other two groups, and (2) to recruit participants of legal age, similar to those included in previous research (in Colombia it is common to start university at 17 years of age). The mean age of the university students was 21.85 years ($SD=1.70$, 31 female). For urban and rural participants, the requisites were people between 30 and 55 years old¹ with a low educational level (as a maximum, they could have completed the compulsory education in Colombia, which finishes in the ninth grade at the age of 14–15 years). To account for inter-area mobility, we also set the requisite that urban and rural populations must have been living in the area for a minimum of 10 years. Urban participants were 43.84 years old ($SD=8.36$, 35 female) and lived in the area for an average of 33 years ($SD=13.73$). They studied on average until sixth grade (11–12 years old) and 27% had completed compulsory education. Rural participants were 34.25 years old ($SD=5.73$, 38 female) and lived in the area for an average of 13 years ($SD=2.85$). They studied on average until seventh grade (12–13 years old) and 13% had completed compulsory education.

University students completed the experiment in Bogotá, the largest city in Colombia; urban participants lived in different neighbourhoods in Medellín, the second-largest city in Colombia; and rural participants lived in the *vereda* Loma Verde. A *vereda* is a Colombian administrative territorial subdivision for rural areas. *Veredas* may include a very small urban centre with two or three streets and a few one- or two-storey buildings. Most of the houses and population are scattered along a large territory linked with dirt roads. To better grasp the difference between urban and rural areas in Colombia, we uploaded pictures of the places in which data collection took place to the Open Science Framework (OSF) website of the project.

¹ The Colombian government defines the age limit for the start of “old age” at 60 years, but decreases it to 55 years if there is physical, vital, or psychological decline (https://www.mininterior.gov.co/sites/default/files/7._enfoco_diferencial_para_personas_mayores.pdf, retrieved 21 February, 2022).

Materials and procedure

We used the video of the film *The stick-up* (Herrington, 2002) also used by Luna and Martín-Luengo (2012). Their results provide an interesting indirect comparison from university students in a WEIRD country. The 3-min video shows two security guards unloading sacks of money into a safe deposit room and walking away. Then, an armed robber in disguise enters the bank, threatens customers and clients, grabs the money, and runs away in a getaway car. The audio track from the video was in Spanish from Spain, which slightly differs from Colombian Spanish. Thus, the video was played without audio to avoid distracting participants with a foreign accent (a similar measure was used in Luna et al., 2015). Despite not having an audio track, the video was still easy to follow. We also used the set of 40 questions by Luna and Martín-Luengo (2012), adapted to the local variant of Spanish. We removed six questions that referred to oral interchanges and used the remaining 34 questions.

We contacted participants through a mix of convenience sampling (i.e., approaching people in the street) and snowball sampling (i.e., a person meeting the requisites would tell us about another person who may be willing to participate). Data collection took place during the COVID-19 pandemic. To minimize the chances of contagion, before starting the experiment participants were given a personal protection kit that included a surgical mask and a small bottle of hydroalcoholic gel. Research assistants also received materials and instructions to protect themselves.

For each participant, the experimenter first introduced himself and explained the requisites of the experiment (e.g., duration and basic tasks). For participants showing interest, the experimenter then asked for permission to audio record the entire exchange. After that, participants answered questions to check the requisites for their group (e.g., for urban and rural populations: age, education level, and years living in the area; for university students: age and number of semesters enrolled at the university). If requisites were met, the experiment moved on. Otherwise, participants were thanked for their time and dismissed.

Participants who were to participate in the experiment then read and signed the consent form and received the protection kit. Then, the experimenter played the video on a 5.5-in. mobile phone screen without audio and with bright at maximum. After the video, participants answered questions regarding their internet exposure. The objective of these questions was twofold. First, they served as a filler task so that the cued recall did not measure short-term memory and, second, they helped us to characterize our participants. The results are summarized in the Online Supplemental Materials available at the OSF website of the project. We did not control the time during the questions and time varied from participant to participant. However, all participants had

3–5 min between the end of the video and the start of the memory test. This time included answering the questions above and reading and explaining the instructions of the memory test.

Finally, the experimenter read aloud each of the 34 questions about the video and participants answered orally to avoid problems with differing levels of reading and writing fluency between participants. Questions could be answered in one word (e.g., “When the robber is seen in the car, what is he holding in his hand?” Correct answer: “A wristwatch”) or in a few words (e.g., “Why did the electricity go out?” Correct answer: “An explosion in an electricity supply pole”). As in Luna and Martín-Luengo (2012), participants were instructed that a “don’t know” answer was not allowed and that they had to provide an answer, even if it was a pure guess. For each answer, participants also reported their confidence that the answer was correct on a scale from 0 (pure guess) to 100 (completely certain that the response was correct) and whether they would like to respond to the question of whether they were witnesses in a trial, with response options of yes or not. A copy of the video, the questions, and the instructions are available on the OSF website in both the original Spanish and translated to English. All the answers were recorded and transcribed after the end of the experiment. Finally, participants were thanked and debriefed about the objectives of the research.

Data analyses

We did not expect differences between groups and thus the popular null-hypothesis significance tests (NHST) were not appropriate because they cannot provide support for the null hypothesis. Instead, we conducted Bayesian analyses and report Bayes factors (*BFs*; for tutorials on Bayesian analyses for psychologists, see Jarosz & Wiley, 2014; Kruschke, 2018; and Wagenmakers et al., 2018).² Bayesian analyses compare two hypotheses and can provide evidence in support of either of them. In the Bayesian analysis of variance (ANOVA) reported below, we compared the hypothesis of no differences between groups (*H*₁) against the hypothesis of differences between groups (*H*₂). For pairwise comparisons, we established a region of proximal equivalence (ROPE) of ± 0.1 standardized units (Kruschke, 2018). The ROPE defines an interval of values that are considered so close to zero that they are assumed to be negligible. By comparing the observed difference between groups against an interval of negligible values, the problems associated with a comparison against a discrete value (i.e., zero) are eliminated (for further discussion, see Kruschke, 2018). The ROPE was defined as 0.1 standardized units because it corresponds to half of what is usually considered a small effect (Cohen’s

² We also conducted analyses based on the NHST and report them in the Online Supplemental Materials. In general, the results from NHST and Bayesian analyses were consistent and led to the same conclusions.

Table 1 Means (standard deviations in parentheses) of the main measures

	University students	Low education urban group	Low education rural group
Proportion Correct Responses	0.53 (0.10)	0.47 (0.14)	0.50 (0.09)
Monitoring Measures			
Confidence Correct Responses	83.62 (9.91)	83.75 (14.21)	85.80 (13.81)
Confidence Incorrect Responses	48.27 (11.57)	57.03 (17.76)	66.72 (20.34)
Gamma Correlation	0.66 (0.20)	0.56 (0.34)	0.52 (0.40)
Area Under the Curve	0.77 (0.09)	0.71 (0.11)	0.66 (0.09)
Control Measures			
Proportion of Responses Reported	0.64 (0.19)	0.81 (0.24)	0.81 (0.20)
Report Criterion (P_{rc})	53.42 (19.03)	32.82 (31.50)	40.87 (29.44)
Memory Benefit Due to the Report Option	0.12 (0.09)	0.07 (0.11)	0.07 (0.09)

$d=0.2$; Kruschke, 2018). For pairwise comparisons, we compared the hypothesis that the average of the difference fell within the ROPE (i.e., $-0.1 < d < 0.1$; H1), meaning no differences or that they are negligible, against the hypothesis that the average of the difference fell outside the ROPE (H2), meaning that differences are not negligible.³ The BF of the comparison would determine the strength of the evidence in support of either hypothesis.

All BF s reported below are BF_{12} and, thus, when higher than 1 they support our hypothesis (H1: no differences between groups), and when lower than 1 they support H2 (there are differences between groups). The further the BF is from 1, the stronger the evidence is in support of either hypothesis. We followed Jeffreys' (1961) recommendations and applied labels to help interpretation, so that BF s between 1 and 3 are labelled anecdotal evidence in support of H1, between 3 and 10 moderate evidence, between 10 and 30 strong evidence, between 30 and 100 very strong evidence, and higher than 100 extreme evidence. Similarly, BF s between 0.33 and 1 are labelled anecdotal evidence in support of H2, and so on with cut-off points of 0.10, 0.03, and 0.01. It is important to note that these cut-off points should not be considered definitive thresholds. $BF=2.90$ and $BF=3.10$ do not provide very different evidence, although they are given different labels. Labels (i.e., anecdotal, moderate...) are only linguistic devices to help interpret and transmit information on the strength of the evidence, and thus we use them liberally here. BF s around 1 are better interpreted as inconclusive, and we arbitrarily determined an interval of inconclusive BF s as those in the range [0.75, 1.25]. Bayesian analyses were conducted with the package BayesFactor (Morey & Rouder, 2018) in R (R Core Team, 2020) and we used the default Cauchy prior ($r=0.707$).

³ There is a more sophisticated version of this approach, which takes into account how much of the posterior distribution falls within the ROPE (see Kruschke, 2018). Here, we opted for the simpler version of the analysis to decide between hypotheses.

Results

Some answers were lost because the recording was unintelligible, the research assistant skipped the question, or the participant failed to provide an answer. This happened for 41 answers for the urban group (2.19% of the answers), 20 answers for the rural group (1.10%), and three answers for the university group (0.16%). In the rural group, we removed the answers from one question because of a procedural error. Unless stated otherwise, we report one-way between-participants Bayesian ANOVAs (group: university students, urban participants, rural participants) followed when appropriate by pairwise Bayesian comparisons between groups. We first present analyses of the proportion of correct responses and then analyses to examine monitoring and control ability. Descriptive statistics are presented in Table 1.

Proportion of correct responses

The three groups watched the video in the street and on a phone screen. It could be argued that those may not be the best viewing conditions. However, when performance was compared with that from Luna and Martín-Luengo (2012), who projected the same video on a large screen in a dim classroom with perfect viewing conditions, our participants showed a similar or higher proportion of correct responses.⁴ Thus, it seems safe to conclude that in the current experiment viewing conditions were satisfactory.

A one-way Bayesian ANOVA showed anecdotal evidence for differences between groups, $BF=0.40$. As the

⁴ Proportion of correct responses for the university students in Luna and Martín-Luengo (2012) was $M=0.43$ ($SD=0.10$). In the three groups reported here, performance was numerically higher, but the difference fell outside the ROPE around 0 for only the university, $BF=4.51 \times 10^{-4}$, and the rural groups, $BF=0.06$, meaning that differences were not negligible. For the urban group there was anecdotal evidence in support of differences falling within the ROPE, $BF=2.22$, suggesting negligible differences.

BF was close to the cut-off for moderate evidence and the corresponding NHST analysis showed significant differences (see the Online Supplemental Materials), we conducted pairwise comparisons to test possible differences between groups. We conducted three analyses that compared the hypothesis that the difference fell within the ROPE (H1) against the hypothesis that the difference fell outside the ROPE (H2). For the comparison between the university and urban groups, the analysis showed moderate evidence in support of H2, $BF = 0.21$. This result indicates that the difference in the proportion of correct responses between the university and urban groups was out of the region around zero or, in simpler terms, that there were differences. The comparison between urban and rural groups showed anecdotal-to-moderate evidence in support of H1, $BF = 2.80$. This result indicates that the difference between groups was so small that it could be safely ignored or, in simpler terms, that there were no differences between groups. The comparison between university and rural groups was inconclusive, $BF = 1.23$. In sum, results suggest that the university group had better memory performance than the urban group.

Monitoring: Resolution measures

Monitoring ability can be studied by checking the degree to which confidence ratings distinguish between correct and incorrect responses, that is, resolution. If participants can monitor their memories and evaluate them as having high or low chances of being correct, then good resolution would show that they can rate these answers with the appropriate level of confidence. We computed three measures of resolution in search of convergent validity: the confidence gap, the Goodman–Kruskal gamma correlation, and the area under the receiving operator characteristics (ROC) curve (see Table 1).

Probably the simplest monitoring measure is the difference between confidence attributed to correct and incorrect responses, which Moritz et al. (2006) called "the confidence gap". The higher the confidence gap, the better the monitoring ability because participants would be rating correct responses with high confidence and incorrect responses with low confidence. To test participants' monitoring ability, we conducted a Bayesian mixed ANOVA 3 (group: students, urban, rural) $\times$ 2 (response: correct, incorrect) with response as a within-participants variable and the average confidence per participant for correct and incorrect responses as measure. The analysis compared four models against the null model of no effects: (1) a model with only group, (2) a model with only response, (3) the additive model with both group and response but without interaction, and (4) the multiplicative model with both variables and the interaction. The last model showed the highest BF , $BF = 4.29 \times 10^{60}$, and

outperformed the second-best model (the additive model) by a factor of 8.25×10^5 , thus providing extreme evidence in support of an effect of both variables and the interaction (see Fig. 1).⁵

To test the main effects and the interaction, pairwise comparisons were conducted using the ROPE as explained above. In the three groups, there was extreme evidence in support of differences between confidence in correct and incorrect responses falling outside the ROPE, university $BF = 5.64 \times 10^{-31}$, urban $BF = 2.87 \times 10^{-12}$, and for rural $BF = 4.17 \times 10^{-6}$, meaning large differences between groups. For correct responses, there was moderate evidence in support of differences between groups falling within the ROPE, meaning that there were no differences or that they were negligible, university versus urban $BF = 7.45$, university versus rural $BF = 4.22$, and urban versus rural $BF = 5.10$. For incorrect responses, the evidence supported that differences between groups fell outside the ROPE, university versus urban $BF = 0.10$, university versus rural $BF = 2.82 \times 10^{-6}$, and urban versus rural $BF = 0.26$. The descriptive analyses showed the highest confidence in incorrect responses in the rural group, then in the urban group, and lower in the university group.

In sum, the analyses of confidence showed that participants in the three groups were able to monitor their memories and rated correct responses with higher confidence than incorrect responses. In addition, the university group monitored their memories better because they rated confidence in the incorrect answers with lower confidence than the other groups.

Gamma correlation is probably the most popular monitoring measure. It is computed from the number of concordant pairs, in which confidence for correct responses is higher than for incorrect responses, and discordant pairs, in which confidence for correct responses is lower than for incorrect responses. Gamma ranges from +1 to -1, with higher numbers meaning better resolution and 0 meaning no resolution. We first compared gamma for each group against the ROPE around 0 to test for monitoring ability. There was extreme evidence in support of the gammas falling outside the ROPE for the three groups, university $BF = 1.83 \times 10^{-29}$, urban $BF = 4.06 \times 10^{-15}$, and rural $BF = 1.81 \times 10^{-11}$. The Bayesian ANOVA 3 (group: students, urban, rural) showed anecdotal evidence in support of no differences between groups, $BF = 1.37$, which was not consistent with the analysis of confidence above. To further explore this discrepancy, we conducted pairwise comparisons. There was moderate evidence in support of differences between urban and rural

⁵ A replication of the analysis in JASP led to the same conclusion because the BF_{inc} for the three effects were large: group $BF_{inc} = 3.51 \times 10^7$, response $BF_{inc} = 4.00 \times 10^{14}$, and the interaction $BF_{inc} = 2.97 \times 10^6$.

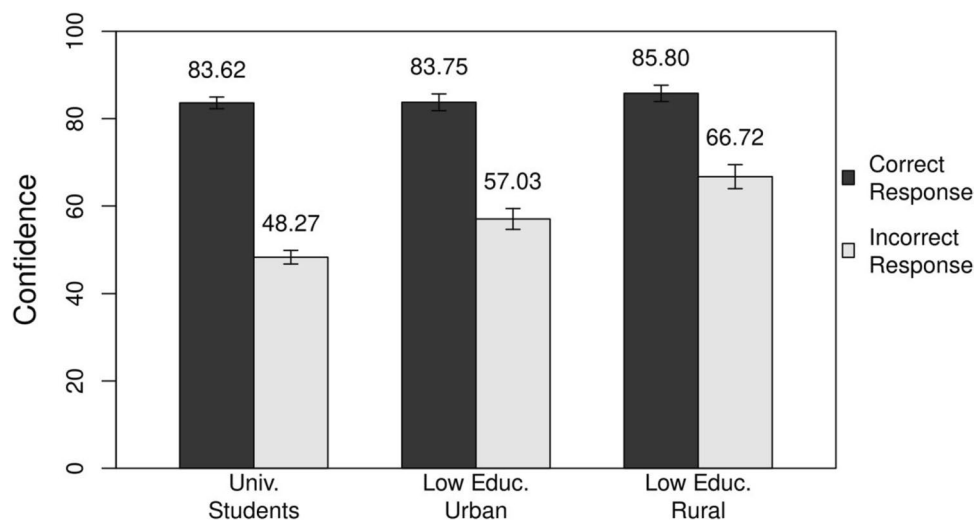


Fig. 1 Mean confidence in correct and incorrect responses per group. Error bars indicate standard error of the mean

groups falling within the ROPE, $BF = 5.99$, anecdotal evidence in support of differences between university and rural groups falling outside the ROPE, $BF = 0.46$, and the comparison between university and urban groups was inconclusive, $BF = 1.01$. In sum, the analyses of gamma showed monitoring in the three groups and hinted that monitoring could be better in the university than in the rural group.

Despite its popularity, gamma has been criticized for having some undesirable properties (Masson & Rotello, 2009). As an alternative, Masson and Rotello (2009) proposed a measure based on the area under the ROC curve (AUC; in the Online Supplemental Materials, see the NHST analyses of AUC for an explanation of computation and meaning, and Fig. S1 for ROC curves). AUC ranges from 0 to 1, with higher numbers indicating better resolution and 0.5 indicating null resolution. We compared AUCs of each group against the ROPE around 0.5. There was extreme evidence in support of AUCs falling outside the ROPE in the three groups, university $BF = 2.88 \times 10^{-28}$, urban $BF = 7.17 \times 10^{-18}$, and rural $BF = 1.14 \times 10^{-15}$. The Bayesian ANOVA showed extreme evidence in support of monitoring differences between groups, $BF = 2.94 \times 10^{-6}$. Pairwise comparisons showed evidence ranging from anecdotal to extreme in support for differences between groups falling outside the ROPE, university versus urban $BF = 0.33$, university versus rural $BF = 1.04 \times 10^{-7}$, and urban versus rural $BF = 0.39$. In sum, AUC showed monitoring in the three groups and that the best monitoring was observed in the university group, followed by the urban group, and then by the rural group.

Finally, to test whether there were differences in monitoring when samples are similar but countries are different, we compared the confidence gap and gammas in our Colombian university students' sample to that of the Spanish students in Luna and Martín-Luengo (2012). The analyses

showed evidence in support of differences falling within the ROPE for both the confidence gap (Colombian $M = 35.35$, $SD = 10.87$ and Spanish $M = 35.24$, $SD = 11.78$), $BF = 7.36$, and gammas (Spanish $M = 0.63$, $SD = 0.17$), $BF = 3.80$. These analyses suggest that there were no monitoring differences between students from a WEIRD and a non-WEIRD country.

In sum, the analyses of the three monitoring measures showed that the three groups could successfully monitor the probability that their memories are correct. Also, results suggest that monitoring is better in the university group than in the other two groups. This difference seems primarily based on differences rating incorrect answers. Results also suggest a lack of differences in monitoring ability when similar samples from different countries were compared.

Control: The report option

To examine the control process, after participants produced an answer we gave them the option to report or withhold that answer were they witnesses in a trial (i.e., the report option). The control process is informed by the output of the monitoring process and good control would happen when participants report correct answers and withhold incorrect answers. We conducted two different sets of analyses to check participants' control of their responses, one based on the proportion of reported answers and another based on the memory benefit that can be achieved via the report option (see Table 1).

For the proportion of responses reported, the Bayesian ANOVA showed extreme evidence in support of differences between groups, $BF = 3.68 \times 10^{-4}$. Pairwise comparisons showed extreme support for differences falling outside the ROPE between university and urban

groups, $BF = 2.99 \times 10^{-3}$, and university and rural groups, $BF = 7.41 \times 10^{-4}$. The university group reported fewer responses than the other two groups. In addition, there was moderate evidence in support of differences falling within the ROPE between urban and rural groups, $BF = 7.40$.

These results suggest that university students may have applied a different confidence criterion to report or withhold answers. Koriatic and Goldsmith (1996) introduced a method to compute that report criterion called report-criterion probability or P_{rc} (for computation details, see also Goldsmith & Koriatic, 2007). A participant's P_{rc} is the level of confidence that better discriminates between reported and withheld answers. If a response is rated with confidence higher than participant's P_{rc} , then it is likely to be reported, and if a response is rated with confidence lower than participant's P_{rc} , it is likely to be withheld. We computed P_{rc} per participant and averaged per group. The Bayesian ANOVA showed very strong evidence in support of differences between groups, $BF = 0.02$. Pairwise comparisons showed evidence in support of differences falling outside of the ROPE between university and urban groups, $BF = 4.99 \times 10^{-3}$, and between university and rural groups, $BF = 0.26$, and anecdotal evidence in support of differences falling within the ROPE between urban and rural groups, $BF = 2.46$. In sum, university students were more conservative and only reported answers for which they had medium-to-high confidence (i.e., higher than 53.42), while urban and rural participants were more liberal and reported answers with lower confidence. These different reporting criteria explain the different proportion of answers reported per group and suggest control differences between groups.

Another way to check the ability to control behaviour is to examine participants' ability to use the report option to increase accuracy. Good control would be shown if participants withhold information with low chances of being correct, resulting in a higher proportion of correct responses for the reported answers when compared with all the answers (i.e., including reported and withheld answers). To measure the memory benefit due to the report option, we computed the difference between the proportion of correct responses for reported answers minus the proportion of correct responses for all the answers (see Table 1). Differences higher than zero would show good control, and the higher the difference, the better the control ability. The memory benefit fell outside the ROPE around 0 for the three groups, university $BF = 2.60 \times 10^{-11}$, urban $BF = 3.30 \times 10^{-3}$, and rural $BF = 1.60 \times 10^{-5}$, thus showing control ability for all participants. We also tested group differences with a Bayesian ANOVA. The results showed moderate evidence in support of differences between groups, $BF = 0.29$. Pairwise comparisons showed that the difference in the memory benefit between the university and urban groups, $BF = 0.33$, and between university and rural groups, $BF = 0.16$, fell outside

the ROPE, and that between urban and rural groups fell within the ROPE, $BF = 7.46$. These results suggest a better control ability in the university than in the other two groups.

In sum, the analyses in this section are consistent in showing that (1) participants in the three groups can control their behaviour using the information from the monitoring process (i.e., confidence), and (2) university students had a better control ability than the other groups.

Discussion

The objective of this research was to study the effectiveness of basic metamemory processes in under-represented samples, particularly in participants with a low educational level from a non-WEIRD country. We expected that the three groups, rural and urban participants with low educational level and a university students control group, would show a functional ability to monitor their memories and to use the input from that process to control their behaviour. The results confirmed that hypothesis, meaning that people from different origins and educational levels can efficiently use their metamemory processes in a task with applied relevance. Also, we expected that educational level would not influence monitoring and control but, instead, we found that these processes were more efficient in university students than in participants with low educational level. We discuss both main results in turn.

The generalizability of psychological findings to all human beings has been challenged because most research is conducted with similar individuals from a limited set of countries (Henrich et al., 2010). Thus, to test the generalizability of psychological phenomena researchers should replicate them across individuals and countries. Our results confirmed that people different from the university students widely used in experimental research, and from a non-WEIRD country, can use the basic metamemory processes in an eyewitness memory task with a reasonable level of success. This is relevant because it should not be taken for granted that metamemory works in all circumstances and types of people. In sum, this research suggests that the basic metamemory processes are functional in participants with different characteristics.

Our results also showed a remarkable similarity between the monitoring ability of university students in a WEIRD country (Spain, in Luna & Martín-Luengo, 2012) and in a non-WEIRD country (Colombia). It does not seem that there are differences across countries if the same type of individuals is used. Instead, it seems like there are differences in metamemory across groups of individuals. These findings support the idea that if behavioural scientists are to generalize phenomena and results to other populations, it may be better to first replicate them across different

types of individuals (Peterson, 2001). Thus, we suggest future researchers attempting to test the generalizability of their results to start with non-student samples in their own countries.

There are likely several demographic variables that may affect the efficiency of metamemory processes. Age, for example, is known to affect metamemory, with children having less efficient metamemory because their cognitive system is under development (Moses-Payne et al., 2021; Schneider & Löffler, 2016). In this research, we explored whether the educational level or the living environment would have any impact on metamemory measures. Educational level and living environment could be indicators of a broader and more complex concept: socio-economic status. Socio-economic status has drawn researchers' attention as an overriding variable to account for behavioural and neural differences between individuals (for a review, see Farah, 2017). For example, within the memory literature several studies have shown that children's socio-economic status is a predictor of their performance on executive function tasks (St. John et al., 2019; Vrantsidis et al., 2020). Our study constitutes a first step to examine the effect of socio-economic variables on metamemory measures, with the effect of the broader concept of socio-economic status on metamemory yet to be explored.

The other main finding of this research is that university students had better overall metamemory than both groups with lower educational level. This study did not test possible mechanisms by which educational level could affect metamemory functioning. However, below we provide some potential explanations.

First, educational level may have affected metamemory processes directly because schooling provides plentiful opportunities to practice monitoring and control. The experience of university students with memory tests (e.g., exams), the feedback over their performance (i.e., grades), the practice with learning strategies, and the assessment of their own learning may have helped them to develop metamemory and make it more efficient. Second, educational level may have had an indirect effect on metamemory by affecting other processes. For example, our findings could be explained by differences in the ability to engage in hypothetical situations or the motivation to exert cognitive effort in a task alien to participants.⁶ Third, educational level might be just one indicator of socio-economic status. As stated above, socio-economic status is a complex variable that has been linked to differences between individuals at several levels: functional brain correlates, cognitive abilities, and physical and

mental health (Farah, 2017). Hence, the differences in metamemory associated with different educational levels could be telling us just a part of a larger story that remains to be told. Whether the differences in memory and metamemory between groups reflect actual differences in metamemory functioning or are due to other processes mediated by or related to education is a matter to be disentangled in future research.

In addition, a relevant issue to understand group differences in this research is that memory and metamemory are related. When memory is better, metamemory is also better (Perfect & Stollery, 1993). Also, there is a peak in memory performance in the early twenties and a slightly decline from there (Murre et al., 2013). Thus, age differences between groups could explain the observed differences in memory, and thus, in metamemory. At a descriptive level, the proportions of correct responses for the three groups is consistent with Murre et al. (2013): higher in the group in their twenties (university students), then slightly lower for the rural (in their thirties), and then in the urban (in their forties). However, the small differences in memory do not seem consistent with the clear lack of differences between urban and rural groups in metamemory. If metamemory differences were due to memory differences, we would have expected a similar pattern to that observed for memory, even if only at a descriptive level. However, there is no such descriptive pattern in metamemory measures. Hence, it seems that educational level might have a stronger effect in metamemory than the memory decline from young to middle adulthood.

Our results also have relevance to eyewitness memory research. Past studies have shown that, under certain conditions, mock witnesses' confidence is highly informative of the accuracy of the memory of what happened during a criminal event (Luna & Martín-Luengo, 2012) or the culprit's identification in a lineup (Wixted & Wells, 2017). However, there are many conditions in which metamemory working is suboptimal (see Wixted & Wells, 2017). Our research showed that the monitoring and control processes needed to rate confidence and decide whether a piece of information is worth reporting or not are also functional in individuals different from university students from WEIRD countries. This is good news for forensic practitioners because witnesses, victims, and perpetrators may come from different backgrounds and will likely vary in many psychological and socio-demographic dimensions. However, this research also showed that memory and metamemory performance was, in general, not as effective for participants with a low educational level. It is premature to forecast whether that difference would be maintained in a real-life situation because it may depend on its explanation. For example, suppose the less efficient metamemory performance was due to a lower motivation to engage in the task. In that case, performance in a real setting may improve and differences between groups

⁶ It is important to note that school attendance is linked to better performance in some cognitive tests but not to a functional criterion of intelligence, such as real-life problem-solving skills (Ardila et al., 2000).

may disappear in an actual police interview. In addition, we used specific procedures to study monitoring and control, and thus our results may be specific to these procedures. Future research aimed at testing different explanations and with different procedures would shed further light on these issues.

In sum, this research showed the need to extend basic cognitive research to different populations with different characteristics. Although results may confirm the presence of a given phenomenon or process and suggest it may be generalizable, such as the monitoring and control processes that form the basis of our understanding of metamemory, there are differences between groups that could remain largely undetected if researchers focus on convenience samples.

Acknowledgements The authors thank Briseida Sánchez and Luis Fernando Cadavid for logistic support.

Declarations

Conflict of interest The authors have no conflicts of interest to disclose.

References

- Apicella, C., Norenzayan, A., & Henrich, J. (Eds.). (2020). Beyond WEIRD [Special issue]. *Evolution and Human Behavior*, 41(5).
- Ardila, A., Ostrosky-Solis, F., Rosselli, M., & Gómez, C. (2000). Age-related cognitive decline during normal aging: The complex effect of education. *Archives of Clinical Neuropsychology*, 15(6), 495–513.
- Argento, O., Pisani, V., Incerti, C. C., Magistrale, G., Caltagirone, C., & Nocentini, U. (2015). The California Verbal Learning Test-II: Normative data for two Italian alternative forms. *The Clinical Neuropsychologist*, 28(Suppl 1), S42–54.
- Arnold, M. M., Higham, P. A., & Martín-Luengo, B. (2013). A little bias goes a long way: The effects of feedback on the strategic regulation of accuracy on formula-scored tests. *Journal of Experimental Psychology: Applied*, 19, 383–402.
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, 8, 44–56.
- Dunlosky, J., & Tauber, S. K. (Eds.). (2016). *The Oxford handbook of metamemory*. Oxford University Press.
- Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through meta-cognitive monitoring and control. *Applied Cognitive Psychology*, 25(3), 501–508.
- Farah, M. J. (2017). The neuroscience of socioeconomic status: Correlates, causes, and consequences. *Neuron*, 96(1), 56–71.
- Goldsmith, M., & Koriat, A. (2007). The strategic regulation of memory accuracy and informativeness. In A. S. Benjamin & B. H. Ross (Eds.), *The Psychology of learning and motivation*, Vol 48: *Skill and strategy in memory use* (pp. 1–60). Academic Press.
- Guerrero-Sastoque, L., Bouazzaoui, B., Burger, L., & Taconnat, L. (2021). Educational level effect on episodic memory performance in older adults: Mediating role of metamemory. *Psychologie Française*, 66(2), 111–126.
- Henrich, J., Heine, S., & Norenzayan, A. (2010). The WEIRD people in the world. *Behavioral and Brain Sciences*, 33(2), 1–75.
- Herrington, R. (Writer/Director). (2002). *The stick-up* [Film]. Universal Pictures Video.
- Izaute, M., & Bacon, E. (2016). Metamemory in psychopathology. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 473–488). Oxford University Press.
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, 7(1), 2–9.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford University Press.
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory. *Psychological Review*, 103, 490–517.
- Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, 1(2), 270–280.
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition*, 43(1), 85–98.
- Loftus, E. F., Donders, K., Hoffman, H. G., & Schooler, J. W. (1989). Creating new memories that are quickly accessed and confidently held. *Memory & Cognition*, 17, 607–616.
- Luna, K. (2019). If it is easy to remember, then it is not secure: Metacognitive beliefs affect password selection. *Applied Cognitive Psychology*, 33(5), 744–758.
- Luna, K., & Martín-Luengo, B. (2012). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology*, 26, 289–295.
- Luna, K., & Martín-Luengo, B. (2014). The subjective experience of retrieval-induced forgetting. *Memory*, 22, 813–823.
- Luna, K., Higham, P. A., & Martín-Luengo, B. (2011). Regulation of memory accuracy with multiple answers: The plurality option. *Journal of Experimental Psychology: Applied*, 17, 148–158.
- Luna, K., Martín-Luengo, B., & Brewer, N. (2015). Are regulatory strategies necessary in the regulation of accuracy? The effect of direct-access answers. *Memory & Cognition*, 43(8), 1180–1192.
- Luna, K., Albuquerque, P. B., & Martín-Luengo, B. (2019). Cognitive load eliminates the effect of perceptual information on judgments of learning with sentences. *Memory & Cognition*, 47, 106–116.
- Masson, M. E. J., & Rotello, C. M. (2009). Sources of bias in the Goodman-Kruskal gamma coefficient measure of association: Implications for studies of metacognitive processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 509–527.
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for common designs. R package version 0.9.12–4.2 [Computer software]. <https://CRAN.R-project.org/package=BayesFactor>
- Moritz, S., & Woodward, T. S. (2006). The contribution of metamemory deficits to schizophrenia. *Journal of Abnormal Psychology*, 115, 15–25.
- Moritz, S., Woodward, T. S., & Chen, E. (2006). Investigation of metamemory dysfunctions in first-episode schizophrenia. *Schizophrenia Research*, 81, 247–252.
- Moses-Payne, M. E., Habicht, J., Bowler, A., Steinbeis, N., & Hauser, T. U. (2021). I know better! Emerging metacognition allows adolescents to ignore false advice. *Developmental Science*, 24(5), e13101.
- Murre, J. M. J., Janssen, S. M. J., Rouw, R., & Meeter, M. (2013). The rise and fall of immediate and delayed memory for verbal and visuospatial information from late childhood to late adulthood. *Acta Psychologica*, 142(1), 96–107.

- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 26, pp. 125–169). Academic Press.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition? In J. M. P. Shimamura (Ed.), *Metacognition* (pp. 1–26). The MIT Press.
- Peng, Y., & Tullis, J. G. (2021). Dividing attention impairs metacognitive control more than monitoring. *Psychonomic Bulletin & Review*, 28(6), 2064–2074.
- Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: One deficit or two? *The Quarterly Journal of Experimental Psychology Section A*, 46(1), 119–135.
- Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78(1), 144–147.
- Perfect, T. J., Hollins, T. S., & Hunt, A. L. R. (2000). Practice and feedback effects on the confidence–accuracy relation in eyewitness memory. *Memory*, 8(4), 235–244.
- Peterson, R. A. (2001). On the use of college students in social science research: Insights from a second-order meta-analysis. *Journal of Consumer Research*, 28(3), 450–461.
- Quattropani, M., Lenzo, V., Faraone, C., Pistorino, G., Bella, I. D., & Mucciardi, M. (2016). The role of metacognition in eating behavior: An exploratory study. *Mediterranean Journal of Clinical Psychology*, 4(3), 1–15.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing [Computer software]. <https://www.r-project.org/>.
- Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences*, 115(45), 11401–11405.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137, 615–625.
- Rosselli, M., & Ardila, A. (2003). The impact of culture and education on non-verbal neuropsychological measurements: A critical review. *Brain and Cognition*, 52(3), 326–333.
- Sauer, J. D., Palmer, M. A., & Brewer, N. (2019). Pitfalls in using eyewitness confidence to diagnose the accuracy of an individual identification decision. *Psychology, Public Policy, and Law*, 25(3), 147–165.
- Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 491–518). Oxford University Press.
- Soderstrom, N. C., Yue, C. L., & Bjork, E. L. (2016). Metamemory and education. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 197–215). Oxford University Press.
- Soler, M. J., & Ruiz, J. C. (1996). The spontaneous use of memory aids at different educational levels. *Applied Cognitive Psychology*, 10(1), 41–51.
- Sporer, S. L., Penrod, S., Read, J. D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence–accuracy relation in eyewitness identification studies. *Psychological Bulletin*, 118, 315–327.
- St. John, A. M., Kibbe, M., & Tarullo, A. R. (2019). A systematic assessment of socioeconomic status and executive functioning in early childhood. *Journal of Experimental Child Psychology*, 178, 352–368.
- Stratta, P., Prosperini, P., Daneluzzo, E., Bustini, M., & Rossi, A. (2001). Educational level and age influence spatial working memory and Wisconsin Card Sorting Test performance differently: A controlled study in schizophrenic patients. *Psychiatry Research*, 102(1), 39–48.
- Szajer, J., & Murphy, C. (2013). Education level predicts retrospective metamemory accuracy in healthy aging and Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, 35(9), 971–982.
- Tiokhin, L., Hackman, J., Munira, S., Jesmin, K., & Hruschka, D. (2019). Generalizability is not optional: Insights from a cross-cultural study of social discounting. *Royal Society Open Science*, 6(2), 181386.
- Vrantsidis, D. M., Clark, C. A. C., Chevalier, N., Espy, K. A., & Wiebe, S. A. (2020). Socioeconomic status and executive function in early childhood: Exploring proximal mechanisms. *Developmental Science*, 23(3), e12917.
- Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Selker, R., Gronau, Q. F., Dropmann, D., Boutin, B., Meerhoff, F., Knight, P., Raj, A., van Kesteren, E.-J., van Doorn, J., Šmíra, M., Epskamp, S., Etz, A., Matzke, D., & Morey, R. D. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25(1), 58–76.
- Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest*, 18(1), 10–65.
- Zarantonello, L., Schiff, S., Amodio, P., & Bisiacchi, P. (2020). The effect of age, educational level, gender and cognitive reserve on visuospatial working memory performance across adult life span. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, 27(2), 302–319.

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

Open practices statement The data, experimental materials, and Online Supplemental Materials for this experiment are available at: https://osf.io/j5v9t/?view_only=7e131bfca6194bc2ac978a768650dc a3