DE GRUYTER
OPEN

Psicológica (2018), 39, 1-24
doi: 10.2478/psicolj-2018-0001

The (limited) effect of emotional arousal in the regulation of accuracy in eyewitness memory

Karlos Luna[*1] & Beatriz Martín-Luengo[2]

[1]*Psychology Research Centre, School of Psychology, University of Minho (Portugal)*

[2]*Centre for Cognition and Decision Making, National Research University-Higher School of Economics (Russian Federation)*

Witnesses encoding a crime are likely to feel negative emotions with high arousal, e.g., anxiety or fear. Negative emotions improve memory for central information and impair memory for peripheral information. In this study we explored the effects of emotional arousal and type of information in the regulation of accuracy. The regulation of accuracy allows participants to maximize accuracy, for example, by deciding on the number of alternatives in their response (the plurality option). Participants were induced with high- and low-arousal negative emotions and then shown a slideshow of a crime. Afterwards, they answered questions about central and peripheral contents of the event. Questions followed the basic plurality option procedure. First, participants selected one alternative (single answer); second, they selected three alternatives (plural answer); and, finally, they decided on reporting either the single or the plural answer. Results showed successful manipulation of arousal, and that the regulation of accuracy led to a greater increase in accuracy for peripheral than for central information, but no differences depending on the level of arousal. We also identified two factors that increased accuracy in the plurality option: the ability to discard answers with low chances of being correct and the addition of answers with higher chances of being correct. Either one, or both, can increase witness accuracy.

[*] **Ackownledgments:** This work was partially supported by the Russian Academic Excellence Project '5-100'. **Corresponding autor:** Karlos Luna. University of Minho. Psychology Research Centre, School of Psychology. Campus de Gualtar, 4710-057, Braga (Portugal). E-mail: karlos.luna.ortega@gmail.com. **Author´s note:** This article was written while the first author was a visiting researcher at the Centre for Cognition and Decision Making, National Research University-Higher School of Economics (Russian Federation). The authors declare that they have no conflict of interest regarding this research.

When someone witnesses a criminal event, it is likely that he or she will feel a negative emotion. Negative emotions include high-arousal emotions, such as anger, and low arousal emotions, such as sadness, but it is unlikely that a witness will feel sadness during a crime. Negative emotions with high arousal, such as distress, anxiety, anger, or fear, are much more likely to be experienced[1]. In addition, when witnesses try to recount a criminal event to police, juries, or lawyers, their confidence that the answer provided is correct is also relevant (Leippe & Eisenstadt, 2007). Confidence is not always diagnostic of memory accuracy (Brewer, 2006), but a growing body of research has shown the usefulness of confidence measures in eyewitness memory, particularly as cues for memory accuracy (e.g., Brewer, Weber, Wootton, & Lindsay, 2012; Luna & Martín-Luengo, 2012a). For example, confidence can be used to regulate the accuracy of an answer by withholding answers rated with low confidence (Koriat & Goldsmith, 1996; Higham, 2007), thus resulting in increased accuracy of the final report (Higham, Luna, & Bloomfield, 2011). The objective of this research was to study the effect of arousal in a negative emotion in the regulation of accuracy for central and peripheral contents. First, we briefly review the relevant findings regarding the effect of arousal and the type of information. We then provide a detailed explanation of the regulation of accuracy, a process that allows participants to maximize the accuracy of their report by relying on their metamemory judgments.

Emotional Arousal. The effect in memory of high-arousal negative emotions, such as those likely to be elicited when witnessing a crime, has been extensively studied, and there is a wealth of research showing that they do affect memory (for reviews, see Kensinger, 2009, and Mather & Sutherland, 2011). Eyewitness memory is not an exception (for a review, see Edelstein, Weede Alexander, Goodman, & Newton, 2004). The consensus among researchers is that negative emotions, when compared against a neutral condition with no emotion, cause an attentional narrowing that will enhance memory for the most relevant information (i.e., central information, see next section), at the expense of peripheral information (Christianson, 1992; for a detailed explanation, see Deffenbacher, Bornstein, Penrod, & McGorty, 2004). However, high-arousal negative emotions, such as anxiety, may have more general effects. In a recent meta-analysis, Moran (2016) found that anxiety negatively affects working memory capacity in a variety of paradigms and conditions.

---

[1] We do not suggest that all crimes will elicit negative emotions or high arousal. Sometimes a witness may not even notice that there was being a crime. However, in serious crimes that, for example, involve the use of force (e.g., an assault, or a fight) it is likely that witnesses will feel a negative emotion with high arousal.

Some studies on the effect of emotions and arousal have collected metacognitive measures (e.g., Beaupré & Hess, 2006; Rimmele, Davachi, & Phelps, 2012; Sharot & Yonelinas, 2008). Most of them show that emotion enhances the subjective experience that accompanies a memory (Phelps & Sharot, 2008). For example, taboo words with highly negative emotional semantic content (e.g., fuck, shit, or dick) are rated with higher confidence than neutral words (MacKay et al., 2004). Similarly, Talarico and Rubin (2003) studied the memory of the September 11, 2011 terrorist attack, a highly negative and emotionally arousing event, and concluded that its main characteristic was the high confidence that memories were correct, even when they were not.

In sum, research suggests that high-arousal negative emotions enhance memory for the most relevant information, i.e., central items, and confidence for all types of items. The effect of arousal in a negative emotion in the regulation of accuracy has not been investigated before. Will the regulation of accuracy and its benefits be different depending on the witness's arousal?

Central and Peripheral Information. A general definition of item centrality usually refers to the importance of an item in an event (Wyler & Oswald, 2016). A more specific definition is that central information is directly linked to the main actions, characters, and objects of the event, and peripheral information is irrelevant (Burke, Heuer, & Reisberg, 1992; Christianson, 1992). In eyewitness memory the distinction is important because central contents are likely to be more relevant in a police investigation. Examples of central information with forensic interest are the weapons used during the crime, the clothing, descriptions, and actions of the criminals, and the reactions of the victims. Peripheral information includes background details which do not play a role in the unfolding of the events or are forensically irrelevant.

Research has shown that memory is better for central than for peripheral information (Lowe, Takarangi, Humphries, & Wright, 2016; Wyler & Oswald, 2016), and that confidence ratings are higher for central than for peripheral contents (Luna & Migueles, 2009; Migueles & García-Bajos, 1999). Since no studies to date have examined the regulation of accuracy with central and peripheral information, the experiment reported here should be considered exploratory. Will the memory benefits of the regulation of accuracy be different for central and peripheral items?

The Regulation of Accuracy. The regulation of accuracy is a process that allows accuracy to be maximized (Luna, Martín-Luengo, & Brewer, 2015). It is based on the fact that people can distinguish answers very likely to be correct from answers very unlikely to be correct, and that they do so using confidence ratings as proxy for accuracy (for a review, see Goldsmith, 2016). For example, if a witness is asked how the bank robber concealed his face, she can answer with a single unit of information, e.g. “with a mask” (single answer), or with several, e.g., “with a mask or with a balaclava” (plural answer). If the witness considers that the single answer has low chances of being correct, she may prefer to report the plural answer, which has higher chances of being correct, in an attempt to increase accuracy. By reporting single answers when they are likely to be correct and plural answers when not, the accuracy of the final report increases. This increase is termed the *memory benefit* of the regulation of accuracy. The regulation of accuracy by adding alternatives is known as the plurality option (Luna, Higham, & Martín-Luengo, 2011), and has proven successful in increasing accuracy (Higham, 2013; Luna et al., 2011; see also the grain size regulation of accuracy for research based on the same principles, e.g., Ackerman & Goldsmith, 2008; Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Goldsmith, Koriat, & Pansky, 2005; Weber & Brewer, 2008).

Of particular interest here is the regulation of accuracy in eyewitness memory, because it provides a straightforward way to increase accuracy, although at the expense of the informativeness of the answer. For example, Luna et al. (2011) found that the accuracy for all single answers, the equivalent of the accuracy in a multiple-choice test with forced answer, was .41 (Experiment 1) and .28 (Experiment 2), and that the accuracy of the final report, which included a mix of single and plural answers, increased to .66 and .59, respectively. The power of the regulation of accuracy was shown in a study examining the effect of misinformation (Luna & Martín-Luengo, 2012b). Even when false information was presented, accuracy increased from all single answers (.36) to the final report (.57).

To date, only a few variables of interest in eyewitness memory have been tested in relation to the regulation of accuracy (for example, the aforementioned study on misinformation). Here we tried to fill this gap, albeit partially, by studying the effects in the regulation of accuracy of two variables that are known to affect eyewitness memory: arousal and type of information. Specifically, we predicted that if high arousal increases confidence, then the proportion of single answers should also be greater with high than with low arousal. The logic is the same for the type of information. If confidence is higher for central items, then we expect more single answers for central than for peripheral information.

The Present Research. Our interest here was to study the effect of arousal in the regulation of accuracy in a criminal event. Ideally, we could have manipulated the arousal induced by the criminal event itself, for example, by comparing memory for an arousing and a neutral event. The problem is that this way we would be testing memory for different events. Loftus, Loftus, and Messo (1987) presented an emotional video of a customer pointing a gun at the cashier of a restaurant, and a neutral video in which the customer handed a check to the cashier. Even though customer, cashier, location, and actions were the same, the event was completely different (robbing vs. paying). Instead, we preferred to test memory for the same event under different arousal conditions. This way, we can be certain that any difference in the regulation of accuracy is due to arousal and not to different materials. For this reason, we resorted to an external manipulation to induce arousal. As crimes likely elicit negative emotions, we tried to ensure that participants were feeling a negative emotion and that the arousal difference between the high- and low-arousal groups was as great as possible. To this end, we employed an arousal induction protocol to put participants in the emotional state that better corresponded to a high-arousal negative emotion (e.g., anxiousness) and a low-arousal negative emotion (e.g., sadness). Witnesses are unlikely to feel sadness when they watch a crime, so here this condition was included as a baseline to test the effect of high arousal.

To induce the relevant emotional state we asked participants to write down a personal experience while listening to music congruent with their condition. They then watched a slideshow about a theft and, after a few minutes, completed a memory test that followed the standard plurality option procedure and included questions about central and peripheral items.

METHOD

Participants. As the main objective was to test the effects of our manipulations on the regulation of accuracy with the plurality option, we determined sample size from past studies on the plurality option. The first study using the plurality option, Luna et al. (2011), reported *Cohen's d* of 1.37 and 0.70, and other studies found similar effect sizes (e.g., Luna et al., 2015). Power analysis suggested that the minimum sample size to find an effect size of 0.70 was 18 participants, with $\alpha = .05$ and $1 - \beta = .80$. Fifty-four university students (37 females, mean age 20.49 years, $SD = 6.42$) started the experiment. Ten participants were dismissed after they completed the Beck Depression Inventory because their score was higher than 12 (see below), and data from one participant was lost due to human error. Thus, 43 participants completed the experiment (28 females, mean age 19.49 years, $SD = 3.69$). Participants were randomly assigned to groups.

Design. A mixed factorial design 2 (arousal: high, low) x 2 (type of item: central, peripheral) was used, with arousal manipulated between subjects and type of item within subject. Twenty-one participants were randomly assigned to the high-arousal group and 22 to the low arousal group.

Materials.

Beck Depression Inventory (BDI). The BDI (Beck, Ward, Mendelsohn, Mock, & Erbaugh, 1961) is a 21-question test for measuring depression. We used it to identify participants with depression or a near-depression disorder who could be negatively affected by the arousal manipulation. Kendall, Hollon, Beck, Hammen, and Ingram (1987) suggested that scores higher than 17 were to be considered a depressive state, but Vaz Serra and Pio Abreu (1973), in the Portuguese adaptation of the BDI, suggested a cut-off score of 12, which was applied here.

Arousal induction. To create the two experimental conditions of negative emotion with high and low arousal we adapted two previously used procedures: (1) Jefferies, Smilek, Eich, and Enns (2008) asked participants to recall in detail an event from their past in which they felt a given emotion while listening to music that fit the mood (i.e., think about a sad event and listen to sad music to induce a sad mood), and (2) Gasper and Clore (2002) asked participants to write about a personal life event that made them feel happy or sad.

To induce high and low arousal we asked participants to write for ten minutes about a personal event that made them feel either anxious (for the high-arousal condition) or sad (for the low-arousal condition), while listening to music congruent with the mood. Anxiety and sadness were selected because they are representative emotions with negative valence and, respectively, high and low arousal. For the musical pieces, we chose a selection of music suggested by Jefferies et al. (2008)[2]. Participants in the high-arousal condition listened to Vivaldi's Four Seasons, Concerto 1: Spring, Presto, and Beethoven's Symphony No. 6 in F Major, Opus 68, Pastoral IV. Participants in the low arousal condition listened to Sibelius' Violin Concerto: Adagio di Molto. The last piece was shorter than ten minutes (7'25") so we put it in a ten-minute loop.

---

[2] The musical pieces used by Jefferies et al. (2008) are listed at http://visionlab.psych.ubc.ca/research/mood-induction-procedures/ (retrieved the 27th March, 2017).

Affect grid. The affect grid (Russell, Weiss, & Mendelsohn, 1989) is a single-item scale that measures both the arousal and pleasantness (or valence) dimensions. It consists of a 9x9 grid with the labels "extremely high arousal" at the top, "extremely low arousal" at the bottom, "extremely pleasant feelings" on the right, and "extremely unpleasant feelings" on the left. The instructions include definitions of both arousal and pleasantness with examples. Participants are required to mark one of the cells in the grid, thus providing both arousal and pleasantness scores with a single answer.

Robbery slideshow and questions. We used a slideshow from Luna and Migueles (2005, also used in Luna et al., 2011, Exp.1) that included 21 slides about a minor theft on a university campus. This slideshow depicted a couple coming out of a building and sitting on the grass. A young man approaches them to ask for a lighter and steals a mobile phone. After a brief scuffle between the thief and one of the friends, the thief runs away. The slideshow did not include audio but all the actions were clear from gestures and body positions. For example, the action of asking for a lighter was easy to understand because the thief had a cigarette in his hand, which he extended towards the couple. The same 20 questions with five alternatives from Luna et al. (2011, Exp.1) were used here (see Appendix). The authors, both of whom have expertise in the evaluation of central and peripheral contents, rated the centrality of the questions based on the definitions of Burke et al. (1992) and Christianson (1992). Central information was defined as any feature perceptually or conceptually related with the main actions and characters of the event, and peripheral information as that irrelevant or spatially distant from the main action and characters. Both authors agreed that eight questions were about central contents and ten about peripheral contents. There was not agreement on the classification of two questions, so they were not included in the analyses.

In the test phase, for each question participants had to provide several answers following the standard plurality option procedure. Below the questions, five alternatives were presented preceded by a number. Participants were asked to first write down the number of their preferred alternative in a text box (single answer) and then to rate the confidence that it was correct on a scale from 0 (not certain at all) to 100 (totally certain) in deciles. Secondly, participants were asked to write down the numbers of three alternatives in three text boxes (plural answer) and to rate the confidence that the correct alternative was one of the three. Twenty participants answered the single answer first and then the plural answer, and the rest answered the plural first and then the single answer. Finally, participants indicated whether they would rather report the single or the plural answer if they were in a trial.

Procedure. Participants entered the lab, either individually or in groups of a maximum of four, and sat in front of a computer. After giving consent, they completed a paper-and-pencil version of the BDI without time constraints. Participants that did not pass the cut-off score were dismissed at this point. The remaining participants then completed the affect grid for the first time (Grid 1). After that, they went through the arousal induction for 10 minutes, and immediately afterwards completed the affect grid again (Grid 2). Participants had no access to their previous affect grid scores.

The participants then watched the slideshow about the robbery. Slides were presented for two seconds each, with a one-second blank screen between slides. After a new affect grid (Grid 3), they completed a filler task for seven minutes consisting of several word searches. After another affect grid (Grid 4), participants answered the questions about the slides at their own pace. After a final grid (Grid 5), participants watched a humorous video clip about a toddler. The objective of this video was to remove any negative affect that could have remained from the arousal induction phase.

RESULTS

To test our hypotheses we conducted analyses of variance (ANOVA) and Student's *t* test. We report partial eta squared (η_p^2) as measure of effect size for the ANOVAs, and Cohen's *d* (*d*) for Student's *t* tests. All reported confidence intervals (*CIs*) are 95 percent confidence intervals.

Arousal induction. The main scores of all the affect grids are presented in Figure 1. Our primary objective with the arousal induction phase was to create two groups, one with high arousal and another with low arousal, during the encoding of the criminal event. Five comparisons were conducted, resulting in a Bonferroni corrected alpha of .010. In the first measure (Grid 1) there were no differences in arousal between the high- and low-arousal groups, $t(41) = 1.26$, $p = .214$, $d = 0.40$. That is, participants in both conditions came to the laboratory with similar levels of arousal. After arousal induction (Grid 2), the score was higher for the high-arousal than for the low-arousal group, $t(41) = 2.90$, $p = .006$, $d = 0.90$. After the slideshow (Grid 3), the difference was still reliable, $t(41) = 2.70$, $p = .010$, $d = 0.84$. The arousal scores for Grid 2 and 3 confirmed that arousal was higher for the high-arousal than for the low-arousal group for the duration of the slides. After the seven-minute filler task, and just before the questions, differences between groups were no longer significant (Grid 4), $t(41) = 1.31$, $p = .197$, $d = 0.42$. After the questions both groups also had similar arousal (Grid 5), $t(41) = 0.54$, $p = .592$, $d = 0.17$. The scores for the last two grids (Grids 4 and 5) guaranteed that there were no differences in arousal during retrieval. Therefore, any differences between the high- and low-arousal groups in memory and metamemory cannot be attributed to different arousal levels during retrieval and are most likely explained by differences of arousal during encoding.

It is also worth noting that participants came to the lab with relatively high arousal. This may reflect the fact that the lab building is detached from the lecture hall, and students have to walk mostly uphill to get there. Because of the high baseline arousal, it may look as if our manipulation mostly reduced arousal in the low-arousal group. However, once the physiological arousal was gone it is most likely that our manipulation also increased arousal in the high-arousal group. Notwithstanding this limitation, the important point is that both groups showed the expected differences in arousal that make it possible to test the effect of arousal on the regulation of accuracy.

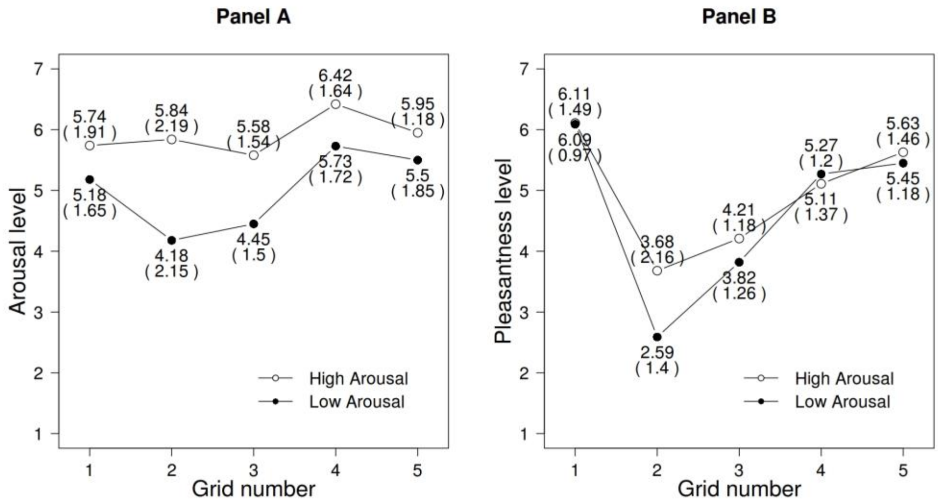

Figure 1. Arousal (Panel A) and pleasantness (Panel B) scores of the five affect grids per group.

The arousal induction was also intended to reduce pleasantness ratings to imitate the negative affect likely experienced by a witness. Since we induced two negative emotions, we expected pleasantness ratings to be similar in both groups. Seven comparisons were conducted, resulting in a Bonferroni corrected alpha of .007. In Grid 1 there were no differences in pleasantness between groups, $t(41) = 0.11$, $p = .911$, $d = 0.04$. After the arousal induction (Grid 2), the pleasantness scores for both high- and low-arousal groups were not significantly different with the adjusted alpha in this section, $t(41) = 2.51$, $p = .016$, $d = 0.78$. However, the high effect size suggests that there may have been some difference in pleasantness between the two groups, which may have affected the results. After the slideshow (Grid 3), there were again no differences between groups, $t(41) = 0.74$, $p = .461$, $d = 0.23$, meaning that any effect of the arousal induction in pleasantness was transient and that it may have affected encoding only during part of the slideshow. We will come back to this issue in the Discussion. In Grids 4 and 5 there was no difference in pleasantness scores between groups, $t(41) = 0.42$, $p = .679$, $d = 0.13$, and $t(41) = 0.41$, $p = .687$, $d = 0.13$, respectively. As intended, the arousal induction decreased the pleasantness ratings from Grid 1 to Grid 2 for both high-arousal, $t(20) = 3.45$, $p = .002$, $d = 0.92$, and low-arousal groups, $t(21) = 11.41$, $p < .001$, $d = 2.90$.

Confidence in Single Answers and Proportion of Single Answers. We expected higher confidence and higher proportion of single answers with central than with peripheral information and with high arousal than with low arousal. To test our predictions we conducted two ANOVAs 2 (arousal: high or low) by type of item (central or peripheral) with confidence in all the single answers and with the proportion of selection of single answers.

For confidence in all single answers, results showed higher confidence for central than for peripheral items ($M = 55.55$, $SD = 19.90$, CI [46.60, 61.50] and $M = 30.95$, $SD = 19.78$, CI [25.04, 36.87]), $F(1, 41) = 125.59$, $p < .001$, $\eta_p^2 = .75$, but no effect of arousal (for high arousal $M = 45.90$, $SD = 22.39$, CI [39.13, 52.67] and for low arousal $M = 40.73$, $SD = 24.08$, CI [33.61, 47.84]), $F(1, 41) = 0.84$, $p = .366$, $\eta_p^2 = .02$. The interaction was not significant, $F(1, 41) = 0.83$, $p = .366$, $\eta_p^2 = .02$.

Results for the proportion of selection of single answers followed the same pattern as those of confidence. Participants selected more single answers for central than for peripheral items ($M = .51$, $SD = .27$, CI [.43, .59] and $M = .24$, $SD = .25$, CI [.16, .31]), $F(1, 41) = 70.82$, $p < .001$, $\eta_p^2 = .63$ but arousal did not affect the proportion of single answers selected ($M = .39$, $SD = .19$, CI [.30, .47] for high arousal and $M = .35$, $SD = .28$, CI [.24, .47] for low arousal), $F(1, 41) = 0.22$, $p = .638$, $\eta_p^2 = .01$. The interaction was not significant, $F(1, 41) = 2.12$, $p = .153$, $\eta_p^2 = .05$.

Regulation of Accuracy. To test whether the regulation of accuracy was successful we compared the accuracy for single selected answers against the accuracy for single rejected answers. A 2 x 2 ANOVA with the difference between accuracy for single selected and single rejected answers was not appropriate because it compares different conditions, and the interest here was to test whether accuracy for single selected was higher than for single rejected answers in each condition. Instead, we conducted four comparisons, one for each condition (adjusted alpha for this section = .013). Accuracy for single selected and single rejected answers per condition is presented in Table 1.

Table 1. Mean (Standard Deviation) [95 percent Confidence Interval] Accuracy for Single Selected and Single Rejected Answers per Condition.

	Single Selected	Single Rejected	Cohen's *d* *
Central information			
High arousal	.54 (.32) [.40, .68]	.32 (.22) [.23, 42]	0.78
Low arousal	.71 (.25) [.59, .84]	.22 (.23) [.11, .33]	2.04
Peripheral information			
High arousal	.19 (.31) [.03, .33]	.20 (.24) [.09, .32]	0.07
Low arousal	.34 (.40) [.13, .54]	.12 (.13) [.06, .19]	0.72

Note. * Cohen's *d* of the difference between single selected and single rejected.

For central information, accuracy for single selected answers was higher than for single rejected answers for both the high-arousal group, $t(20) = 2.92$, $p = .008$, $d = 0.78$, and the low-arousal group, $t(15) = 15.51$, $p < .001$, $d = 2.04$. For peripheral information, accuracy for single selected was marginally higher than for single rejected answers in the low-arousal group, $t(14) = 2.07$, $p = .057$, $d = 0.72$. In interpreting this last result we should note the low n (15) for this analysis, because several participants did not select any single answers, suggesting that the power to detect a real difference may be low. On the other hand, the large effect size suggests that there is a difference. Here we favour a conservative interpretation and prefer to conclude that there were no differences. For the high-arousal group there was no difference in accuracy between single selected and single rejected answers, $t(15) = 0.19$, $p = .851$, $d = 0.06$. In summary, the analyses showed that participants can successfully regulate accuracy for central information, but that for peripheral information, particularly in high-arousal conditions, regulation is not successful.

Memory benefits of the Regulation of Accuracy. The main applied interest of the regulation of accuracy is that when participants are allowed to control their response, the accuracy of the final report increases. We conducted four comparisons to test whether accuracy for the final report was higher than the accuracy for all the single answers, either selected or rejected (adjusted alpha for this section = .013). Accuracy for all single answers and for the final report is presented in Table 2.

Table 2. Mean (Standard Deviation) [95 percent Confidence Interval] Accuracy for All the Single Answers and the Final Report per Condition.

	All Single	Final Report	Cohen's *d* *
Central information			
High arousal	.44 (.22) [.35, .53]	.64 (.20) [.56, .73]	0.97
Low arousal	.42 (.14) [.37, .48]	.70 (.15) [.63, .76]	1.87
Peripheral information			
High arousal	.20 (.14) [.14, .26]	.54 (.19) [.46, .62]	2.11
Low arousal	.20 (.12) [.14, .25]	.59 (.18) [.51, .66]	2.51

Note. * Cohen's *d* of the difference between all single and the final report.

All the comparisons were significant and showed high effect sizes: for central items and high arousal, $t(20) = 6.17$, $p < .001$, $d = 0.97$, for central items and low arousal, $t(21) = 6.10$, $p < .001$, $d = 1.87$, for peripheral items and high arousal, $t(20) = 9.76$, $p < .001$, $d = 2.11$, and for peripheral items and low arousal, $t(21) = 11.25$, $p < .001$, $d = 2.51$. We found that the regulation of accuracy with the plurality option increased accuracy in all four conditions. It is particularly interesting to note that for peripheral items in the high-arousal condition the increase in accuracy was large, even though participants had difficulties distinguishing between a single correct and a single incorrect answer.

To test if there was any difference in memory benefits depending on arousal and type of information, a 2 x 2 ANOVA was conducted on the difference between accuracy for the final report minus accuracy for all single answers. The memory benefit was higher for peripheral ($M = .37$, $SD = .16$, CI [.32, .42]) than for central items ($M = .24$, $SD = .18$, CI [.18, .28]), $F(1, 41) = 16.15$, $p < .001$, $\eta_p^2 = .28$, and there were no differences between high ($M = .28$, $SD = .11$, CI [.23, .32]) and low ($M = .33$, $SD = .15$, CI [.27, .39]) arousal conditions, $F(1, 41) = 1.78$, $p = .189$, $\eta_p^2 = .04$. The interaction was not significant, $F(1, 41) = 0.10$, $p = .751$, $\eta_p^2 = .002$.

The Confidence-Accuracy Relationship. Participants in the high-arousal group were unable to distinguish between answers with high and low chances of being correct for peripheral items. The general framework of the regulation of accuracy states that this ability is affected by both absolute and relative monitoring (Goldsmith & Koriat, 2007). Therefore, we computed absolute and relative monitoring measures to further explore this issue (see Table 3).

Table 3. Mean (Standard Deviation) [95 percent Confidence Interval] of the Relative (Gamma) and Absolute (Over/Underconfidence) Monitoring Measures.

	Gamma	Over/Underconfidence
Central information		
High arousal	.57 (.50) [.35, .80]	.15 (.19) [.07, .23]
Low arousal	.64 (.41) [.46, .81]	.09 (.22) [.002, .19]
Peripheral information		
High arousal	.01 (.61) [-.29, .30]	.13 (.21) [.04, .22]
Low arousal	.26 (.63) [-.06, .58]	.10 (.24) [-.003, .20]

Relative monitoring (also known as resolution) is the extent to which confidence distinguishes correct from incorrect answers and is usually measured with the within-subjects Goodman-Kruskal gamma correlation. For each participant, we computed gamma between accuracy and confidence for all single answers and averaged the results. For central items gammas were higher than zero with both high, $t(18) = 5.04$, $p < .001$, $d = 1.68$, and low arousal, $t(20) = 7.18$, $p < .001$, $d = 2.27$. For peripheral items none of the correlations was different from zero: for high arousal, $t(15) = 0.04$, $p = .971$, $d = 0.01$, and for low arousal, $t(14) = 1.57$, $p = .139$, $d = 0.59$.

An ANOVA 2 arousal x 2 type of information showed better gamma for central ($M = .65$, $SD = .42$, CI [.49, .80]) than for peripheral items ($M = .13$, $SD = .64$, CI [-.10, .36]), $F(1, 27) = 13.77$, $p < .001$, $\eta_p^2 = .34$, and a marginally higher gamma for low ($M = .52$, $SD = .28$, CI [.38, .66]) than for high arousal ($M = .24$, $SD = .45$, CI [.01, .47]), $F(1, 27) = 4.10$, $p = .053$, $\eta_p^2 = .13$. The interaction was not significant, $F(1, 27) = 0.01$, $p = .930$, $\eta_p^2 < .001$. In sum, results for gammas showed the same pattern as the analyses of the difference between accuracy for single selected and single rejected answers, supporting the relevance of resolution in the regulation of accuracy.

Absolute monitoring measures the correspondence between the proportion of correct responses and the confidence ratings. Good absolute monitoring happens when, for example, 80 percent of the answers rated with confidence 80 are correct. A straightforward measure of absolute monitoring is the over/underconfidence (O/U) index, i.e., the difference between the mean confidence and the proportion of correct responses[3]. We computed O/U for single answers for each participant (confidence divided by 100 minus accuracy) and averaged the results per condition. Higher O/U means greater difference between confidence and accuracy and worse absolute monitoring. Positive O/U indicates overconfidence and negative O/U indicates underconfidence.

All four comparisons showed marginal or significant differences from zero and medium-to-large effect sizes, showing overconfidence and suggesting generally poor calibration (for central items and high arousal, $t(20) = 3.67$, $p = .001$, $d = 1.16$, for central items and low arousal, $t(21) = 2.00$, $p = .058$, $d = 0.62$, for peripheral items and high arousal, $t(20) = 2.82$, $p = .010$, $d = 0.89$, and for peripheral items and low arousal, $t(21) = 1.91$, $p = .070$, $d = 0.59$). These results contrast with the high gammas in some conditions. An ANOVA 2 arousal x 2 type of information did not show any effect, lower $p = .451$, $\eta_p^2 = .01$.

DISCUSSION

The main objective of this study was to explore the effect of arousal and the type of information in the regulation of accuracy. We obtained several key and novel results. First, the type of information, but not the level of emotional arousal, affected both confidence and the proportion of single selected answers. Second, participants with high arousal were unable to distinguish answers with high and low chances of being correct for peripheral items. Third, despite this, the regulation of accuracy increased the accuracy of the final report for all conditions.

We expected higher confidence with high arousal and central information. As confidence drives the selection of the single or the plural

[3] Calibration curves are another way to measure absolute monitoring. However, it has been suggested that a minimum of 200 data points per confidence level are necessary for a calibration curve to be reliable (Juslin, Olson, & Winman, 1996). Our experiment did not allow to reach those numbers so we turned to the simpler over/underconfidence measure.

answer, we also expected more single answers for these conditions. Our results supported the effect of the type of information, but not the effect of arousal. In fact, arousal affected none of the main measures[4]. However, the effects of arousal in accuracy and confidence are well established. We offer two tentative explanations for the discrepancy between our results and others that have shown an effect of arousal.

The first explanation is related with the absolute level of arousal. We employed a widely used emotional arousal induction and were successful in creating two groups with high and low arousal during encoding. However, the arousal level in the high-arousal group may have not been high enough as to show the usual effects. In the high-arousal group, the level of arousal just after the induction was 6.14 over a maximum of 9 points, so there was still room for even higher arousal. Future research could try to use other procedures to induce more extreme arousal. If our results were confirmed, they may suggest that regulation of accuracy is impervious to emotional arousal.

A second explanation is related with the pleasantness ratings, because the induction also caused differences in pleasantness between groups. The low-arousal group also showed a more negative valence during at least part of the encoding of the criminal event. At this point it is not clear how differences in valence may have affected encoding and the regulation of accuracy, but it is an interesting venue for future research. Crimes are expected to elicit negative emotions in witnesses, but how negative the emotion is may vary. For example, the theft of a mobile phone may not be perceived nearly as negative as a baby being kidnapping from a pram in front of a helpless mother. The study of the effect of valence in the regulation of accuracy will help determine the best conditions under which the regulation of accuracy can increase the accuracy of a witness' testimony.

Another interesting result is that for peripheral items in the high-arousal condition participants were unable to distinguish between answers with high and low chances of being correct, as shown by the analyses of accuracy for single selected vs. single rejected answers and gamma. The conclusion from these results is that the regulation of accuracy did not work for that condition. However, there was still an increase in accuracy from all the single answers (i.e., the accuracy that would have been obtained in a typical experiment with multiple-choice questions and forced answer), to

---

[4] The only exception was that accuracy for single selected answers was higher with low (M = .49, SD = .27, CI [.36, .62]) than with high arousal (M = .32, SD = .18, CI [.24, .41]), $F(1, 30) = 4.34$, $p = .046$, $\eta_p^2 = .13$.

the final report (i.e., the accuracy obtained when participants are allowed to regulate the number of alternatives in their answer).

This apparent paradox can be explained by the fact that the accuracy increase from all single answers to the final report, i.e., the memory benefit of the regulation of accuracy, comes from two main sources. The first occurs when participants reject single answers with low probabilities of being correct. If a subset of answers with low accuracy is removed, the remaining subset must have higher accuracy (see the report option, Arnold, 2013; Arnold, Chisholm, & Pike, 2016; Arnold, Higham, & Martín-Luengo, 2013; Paulo, Albuquerque, & Bull, 2016). The second occurs with the addition of a subset of plural answers with higher accuracy than the single answers. The memory benefit of the regulation of accuracy may come from one source or the other, or, more likely, both.

Our experiment provided, for the first time, examples of all three situations. For example, for central information in the high-arousal condition the benefit in accuracy came from both sources. As per the first source, there was an increase in accuracy from all single answers to selected single answers (from .44 to .54; p = .037, d = 0.36). As per the second source, participants added a subset of plural answers with higher accuracy, leading to an additional increase (from .54 to .64; p = .038, d = 0.39). However, for peripheral information in the high-arousal group the first source did not increase accuracy (from .20 to .19; p = .748, d = 0.10), but the second source did (from .19 to .54; p < .001, d = 1.27). Conversely, for central information in the low-arousal condition the first source did increase accuracy (from .42 to .71; p < .001, d = 1.17), but the second one did not (from .71 to .70; p = .853, d = 0.04).

From an applied perspective, the source of the accuracy increase may not seem particularly relevant, as long as there is an increase. However, we should bear in mind that the increase in accuracy comes at a cost. The price is paid in informativeness, and is different for each source. In the first source the cost is the lower number of answers that remain after discarding those with low chances of being correct (the same cost as if applying the report option, Koriat & Goldsmith, 1996). In the second source the cost is the informativeness of the added plural answers, which is lower than for single answers (Goldsmith et al., 2000). Depending on the situation, it may be preferable to pay for the accuracy increase in one currency or another.

Our results also suggest that calibration, measured with the O/U statistic, was not a very informative measure. Godsmith & Koriat (2007) included calibration as one of the measures of their quantity-accuracy profile (QAP) to study regulation of accuracy. Surely, calibration is informative about the monitoring process at the theoretical level, although caution in its interpretation has been recently proposed (e.g., Hanczakowski, Zawadzka, Pasek, & Higham 2013). However, from a more applied perspective our research has shown that it does not seem to offer relevant information about the conditions that lead to a higher memory benefit from the regulation of accuracy.

The plurality option has a clear applied advantage over other ways to regulate accuracy. For example, the grain size (e.g., narrowing or widening a numerical interval) can only be applied to quantifiable dimensions, while the plurality option can be used with all types of questions and answers. However, it also has an important limitation. If, for example, a witness answers with three units of information, the units may not be compatible, thus compromising the credibility of the witness (Hollins & Weber, 2016). For example, an answer that the robber concealed his face with "a mask, a stocking, or a balaclava" includes three compatible alternatives because they provide some basic information, in this case, that the entire face of the robber was covered and that any attempt at face identification will be futile. On the other hand, the alternatives "a mask, a scarf, or a bandanna" may be seen as incompatible because the three alternatives do not have much in common, making it difficult to extract useful information. In this case, the witness may be perceived as unreliable.

This is a legitimate constraint of the plurality option, one that extends also to some operationalizations of the grain size (e.g., the grain-size lineup in which witnesses can select several members in a lineup, Horry, Brewer, & Weber, 2016). In our experiment we did our best to make alternatives compatible, but the credibility attached by police and other members of the judicial system to answers that include several alternatives should be investigated. However, we can envisage situations in which the information gathered with a plural answer, even with seemingly incompatible alternatives, may compensate for any hypothetically reduced witness credibility. For example, in the course of a cognitive interview in which one of the instructions is to report everything (Fisher & Geiselman, 1992), a witness saying that the robber concealed his face with a mask, with a stocking, or with a bandanna may be positively regarded because she is correctly following the instructions, and at the same time she is at least also providing some information. It will be the investigator's task to rate the usefulness of the answer and the compatibility of the alternatives. Even in an extreme case, the conclusion that the witness has no idea and is not reliable may also be very useful.

To sum up, our results confirm the usefulness of the regulation of accuracy to increase the accuracy of the final testimony, and identified two sources for this increase. One is based on the ability to discard answers with low chances of being correct and the other is based on the addition of answers with higher chances of being correct. Either or both sources contributed to the increased accuracy of eyewitness testimony for both central and peripheral information.

RESUMEN

El (limitado) efecto de la activación emocional en la regulación de la exactitud de la memoria de testigos. Es muy probable que los testigos de un crimen sientan emociones negativas con un alto grado de activación, por ejemplo, ansiedad o miedo. Las emociones negativas mejoran la memoria de información central y empeoran la memoria de información periférica. En esta investigación estudiamos el efecto de la activación emocional y el tipo de información en la regulación de la exactitud. La regulación de la exactitud permite a los participantes maximizar la exactitud, por ejemplo, decidiendo cuántas alternativas quieren incluir en su respuesta (la opción de pluralidad). Se indujeron en los participantes emociones negativas con un grado de activación alto y bajo, y después se presentó una serie de diapositivas sobre un crimen. Después los participantes respondieron preguntas sobre contenidos centrales y periféricos del crimen. Las preguntas siguieron el procedimiento básico de la opción de pluralidad. Primero los participantes seleccionaron una alternativa (respuesta única), segundo seleccionaron tres alternativas (respuesta plural), y finalmente decidieron si preferían escoger la respuesta única o la plural. Los resultados mostraron que la manipulación de la activación tuvo éxito, y que hubo un mayor aumento en la exactitud con información periférica que central gracias a la regulación de la exactitud, aunque no hubo diferencias en función del nivel de activación. También se identificaron dos factores que aumentan la exactitud en la opción de pluralidad: la capacidad de descartar respuestas con bajas probabilidades de ser correctas y la adición de respuestas con mayores probabilidades de ser correctas. Cada uno de esos factores, o los dos juntos, pueden aumentar la exactitud de los testigos de un crimen.

REFERENCES

Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting–with and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1224-1245. doi:10.1037/a0012938

Arnold, M. M. (2013). Monitoring and meta-metacognition in the own-race bias. *Acta Psychologica*, *144*, 380-389. doi:10.1016/j.actpsy.2013.07.007

Arnold, M. M., Chisholm, L. M., & Prike, T. (2016). No pain no gain: The positive impact of punishment on the strategic regulation of accuracy. *Memory*, *24*, 146-153. doi:10.1080/09658211.2014.990982

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. doi:10.1037/a0034833

Beaupré, M. G., & Hess, U. (2006). An ingroup advantage for confidence in emotion recognition judgments: The moderating effect of familiarity with the expressions of outgroup members. *Personality and Social Psychology Bulletin*, *32*, 16-26. doi:10.1177/0146167205277097

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561-571. doi:10.1001/archpsyc.1961.01710120031004

Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. *Legal and Criminological Psychology, 11*, 3-23. doi:10.1348/135532505X79672

Brewer, N., Weber, N., Wootton, D., & Lindsay, D. S. (2012). Identifying the bad guy in a lineup using a confidence judgments under deadline pressure. *Psychological Science, 23*, 1208-1214. doi:10.1177/0956797612441217

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition*, *20*, 277-290. doi:10.3758/BF03199665

Christianson, S. A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, *112*, 284-309. doi:10.1037/0033-2909.112.2.284

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*, 687-706. doi:10.1007/s10979-004-0565-x

Edelstein, R. S., Weede Alexander, K., Goodman, G. S. & Newton, J.W. (2004). Emotion and eyewitness memory. In D. Reisberg & P. Hertel (Eds.), *Memory and emotion* (pp. 308-346). New York, USA: Oxford University Press. doi:10.1093/acprof:oso/9780195158564.003.0010

Fisher, R. P. & Geiselman, R.E. (1992). *Memory-enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, USA: Charles C. Thomas, Publisher.

Gasper, K., & Clore, G. L. (2002). Attending to the big picture: Mood and global versus local processing of visual information. *Psychological Science*, *1*, 34-40. doi:10.1111/1467-9280.00406

Goldsmith, M. (2016). Metacognitive quality-control processes in memory retrieval and reporting. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 357-385). Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199336746.013.28

Goldsmith, G. & Koriat, A. (2007). The strategic regulation of memory accuracy and informativeness. In A. Benjamin & B. Ross (Eds.), *The Psychology of Learning and Motivation, Vol. 48: Skill and Strategy in Memory Use* (pp. 1-60). San Diego, USA: Elsevier. doi:10.1016/S0079-7421(07)48001-X

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language*, *52*, 505-525. doi:10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, *131*, 73-95. doi:10.1037/0096-3445.131.1.73

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language, 69*, 429-444. doi:10.1016/j.jml.2013.05.003

Higham, P. A. (2007). No special K! A signal detection framework for the strategic regulation of memory accuracy. *Journal of Experimental Psychology: General*, *136*, 1-22. doi:10.1037/0096-3445.136.1.1

Higham, P. A. (2013). Regulating accuracy on university tests with the plurality option. *Learning and Instruction*, *24*, 26-36. doi:10.1016/j.learninstruc.2012.08.001

Higham, P. A., Luna, K., & Bloomfield, J. (2011). Trace-strength and source-monitoring accounts of accuracy and metacognitive resolution in the misinformation paradigm. *Applied Cognitive Psychology*, *25*, 324-335. doi:10.1002/acp.1694

Hollins, T. J., & Weber, N. (2016). Monitoring and regulation of accuracy in eyewitness memory: Time to get some control. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 171-195). Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199336746.013.22

Horry, R., Brewer, N., & Weber, N. (2016). The grain-size lineup: A test of a novel eyewitness identification procedure. *Law and Human Behavior, 40*, 147-158. doi:10.1037/lhb0000166

Jefferies, L. N., Smilek, D., Eich, E., & Enns, J. T. (2008). Emotional valence and arousal interact in attentional control. *Psychological Science*, *19*, 290-295. doi:10.1111/j.1467-9280.2008.02082.x

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1304-1316. doi:10.1037/0278-7393.22.5.1304

Kendall, P. C., Hollon, S. D., Beck, A. T., Hammen, C. L., & Ingram, R. E. (1987). Issues and recommendations regarding use of the Beck Depression Inventory. *Cognitive Therapy and Research*, *11*, 289-299. doi:10.1007/BF01186280

Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, *1*, 99-113. doi:10.1177/1754073908100432

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory. *Psychological Review*, *103*, 490-517. doi:10.1037/0033-295X.103.3.490

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The Handbook of Eyewitness Psychology, Vol.2* (pp. 377-425). Mahwah, USA: Erlbaum. doi:10.4324/9781315805535.ch16

Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior, 11*, 55-62. doi:10.1007/BF01044839

Lowe, H. D., Takarangi, M. K. T., Humphries, J. E., & Wright, D. S. (2016). Alcohol and remembering a hypothetical sexual assault: Can people who were under the influence of alcohol during the event provide accurate testimony? *Memory, 24*, 1042-1061. doi:

Luna, K., & Martín-Luengo, B. (2012a). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology, 26,* 289-295. doi:10.1080/09658211.2015.1064536

Luna, K., & Martín-Luengo, B. (2012b). Improving the accuracy of eyewitnesses in the presence of misinformation with the plurality option. *Applied Cognitive Psychology*, *26*, 687-693. doi:10.1002/acp.2845

Luna, K., & Migueles, M. (2005). Efecto del tipo de información sugerida en el paradigma de la información postevento (Effect of the type of false information in the misinformation paradigm). *Revista de Psicología General y Aplicada*, *58*, 309-321. doi:

Luna, K., & Migueles, M. (2009). Acceptance and confidence of central and peripheral misinformation. *The Spanish Journal of Psychology*, *12*, 405-413. doi:10.1017/S1138741600001797

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. doi:10.1037/a0023276

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*, 1180-1192. doi:10.3758/s13421-015-0534-2

MacKay, D. G., Shafto, M., Taylor, J. K., Marian, D. E., Abrams, L., & Dyer, J. R. (2004). Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. *Memory & Cognition*, *32*, 474-488. doi:10.3758/BF03195840

Mather, M., & Sutherland, M. R. (2011). Arousal-biased competition in perception and memory. *Perspectives on Psychological Science*, *6*, 114-133. doi:10.1177/1745691611400234

Migueles, M., & García-Bajos, E. (1999). Recall, recognition and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology*, *13*, 257-268. doi:10.1002/(SICI)1099-0720(199906)13:3%3C257::AID-ACP566%3E3.0.CO;2-7

Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin, 142*, 831-864. doi:10.1037/bul0000051

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2016). The enhanced cognitive interview: expressions of uncertainty, motivation and its relation with report accuracy. *Psychology, Crime & Law, 22*, 366-381. doi:10.1080/1068316X.2015.1109089

Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science, 17*, 147-152. doi:10.1111/j.1467-8721.2008.00565.x

Rimmele, U., Davachi, L., & Phelps, E. A. (2012). Memory for time and place contributes to enhances confidence in memories for emotional events. *Emotion*, *12*, 834-846. doi:10.1037/a0028003

Russell, J. A., Weiss, A., & Mendelsohn, G. A. (1989). Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57*, 493-502. doi:10.1037/0022-3514.57.3.493

Sharot, T., & Yonelinas, A. P. (2008). Differential time-dependent effects of emotion on recollective experience ans memory for contextual information. *Cognition*, *106*, 538-547. doi:10.1016/j.cognition.2007.03.002

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*, 455-461. doi:10.1111/1467-9280.02453

Vaz Serra, A., & Pio Abreu, J. L. (1973). Aferição dos quadros depressivos. I. Ensaio de aplicação do "Inventário Depressivo de Beck" a uma amostra portuguesa de doentes deprimidos [Adaptation of the depressive symptomatology. I. Practice application of the Beck Depression Inventory on a Portuguese sample of depressed patients]. *Coimbra Médica*, *20*, 623-644.

Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, *14*, 50-60. doi:10.1037/1076-898X.14.1.50

Wyler, H., & Oswald, M. E. (2016). Why misinformation is reported: Evidence from a warning and a source-monitoring task. *Memory, 24*, 1419-1434. doi:10.1080/09658211.2015.1117641

APPENDIX

Questions and alternative answers. The plus symbol (+) indicates the correct answer.

Questions about central contents: 3, 5, 14, 15, 16, 17, 18, and 19.

Questions about peripheral contents: 1, 2, 4, 6, 7, 8, 9, 10, 12, and 13.

1. The two guys came out of a white building. Just above them there is a window. What shape was this window? Square Elliptical Triangular Round + Rectangular	11. When they arrived on the grass, the girl gave money to the boy so that he could buy something in a near by café. In what position was the boy taking the money? Lying down on the grass Standing + Kneeling With only one knee on the grass Sitting down
2. The two guys came out of a white building. What material was the building's façade made of? Metallic panels Glass Bricks Wall tiles + Concrete	12. When the girl gave money to the boy, what was mirrored in the building's façade in the background? The sky A tall tree Another building + A crane A mountain
3. The girl had a folder. What colour was it? Black Red Blue Yellow + Violet	13. When the boy was walking away there was a person sitting down on a bench. What object was close to this bench? A litter pen A streetlight Several flowers A small tree +

	A drinking fountain
4. On both sides of the stairs there was/were: A banister A cement ramp A row of bushes Two streetlights A litter bin +	14. When the boy was walking away and the girl was seated on the grass, what was the girl doing? Looking at him + Tying her hair back Taking off her scarf Leafing through a book Searching for something in her coat
5. The girl wore: A large grey coat with blue trousers A large brown coat with blue trousers A large blue coat with a red scarf A large grey coat with black trousers A large brown coat with black trousers +	15. When the robber approached the girl, what did he do? Say hello Ask for a pen Ask for the time Ask for a cigarette Ask for a lighter +
6. When they went out of the building, there was a fence to the right. What was near by? A locked bicycle + A parked motorbike A child on top of the fence A boy crouched down close to the fence A placard	16. What colour was the robber's jacket?: Black Dark blue + Dark red Dark grey Dark green
7. When they had crossed the road, what passed behind them? A bicycle A lorry A van A car A motorbike +	17. The robber stole: A wallet A calculator A mobile phone + An mp3 player A PDA
8. What shape was the telephone box's roof? A blue cylinder A blue pyramid + A green cube A blue cone A green sphere	18. What footwear was the robber wearing? Shoes Flip-flops Sandals Training shoes + Boots

9. When they were making a telephone call there was a building in the background. How many people were going inside? 5 people 2 people 3 people 1 person + 4 people	19. When the robber ran away, what did the girl do? She stood up and shouted for help She stood up to help her friend + She stayed quiet looking at the robber She stood up and furiously threw something at the robber She stood up and ran after the robber
10. To the left there is a car park. How many cars were parked? 5 cars 1 car 3 cars 2 cars + 4 cars	20. Finally they stayed on the grass drinking a soft drink. What was the brand of soft drink? Seven Up IceTea Coca-Cola + Pepsi Fanta

(Manuscript received: 17 February 2017; accepted: 11 April 2017)