

# Are regulatory strategies necessary in the regulation of accuracy? The effect of direct-access answers

Karlos Luna<sup>1</sup> · Beatriz Martín-Luengo<sup>1,2</sup> · Neil Brewer<sup>3</sup>

Published online: 7 July 2015  
© Psychonomic Society, Inc. 2015

**Abstract** Previous studies have shown that, when people asked to retrieve something from memory have the chance to regulate memory accuracy, the accuracy of their final report increases. Such regulation of accuracy can be made through one of several strategies: the report option, the grain-size option, or the plurality option. However, sometimes an answer can be directly accessed and reported without resorting to such strategies. The direct-access answers are expected to be fast, have high accuracy, and be rated with high probabilities of being correct. Thus, direct-access answers alone could explain the increase of accuracy that has been considered the outcome of regulatory strategies. If so, regulatory strategies may not be needed to explain the previous results. In two experiments, we disentangled the effects of direct-access answers and regulatory strategies in the increase of accuracy. We identified a subset of direct-access answers, and then examined the regulation of accuracy with the plurality option when they were removed. Participants answered questions with six (Exp. 1) or five (Exp. 2) alternatives. Their task was, first, to select as many alternatives as they wanted and, second, to select only two or four alternatives. The results showed that the direct-access answer affected the regulation of accuracy and made it easier. However, the results also showed that regulatory strategies, in this case the plurality option, are needed to explain

why the accuracy of final report increases after successful regulation. This research highlighted the relevance of taking direct-access answers into account in the study of the regulation of accuracy.

**Keywords** Metamemory · Regulation of memory accuracy · Plurality option · Direct-access answers

## Introduction

Sometimes our memory is not as good as we would like. Often this does not matter much (e.g., in a casual conversation), but at other times it can have negative consequences, such as in an exam or when trying to recall the events of a crime that one has witnessed. Fortunately, some studies have identified different ways to increase the accuracy of memory reports, with these involving the use of regulatory strategies when our memory is weak. These studies are based on the idea that memory reports can be regulated. But, what if we can access the answer of a question directly, and regulatory strategies are not needed? How might such answers affect the interpretation of the outcomes of studies about the role of regulatory strategies in the regulation of accuracy? The aim of this research was to examine the impact of answers that do not require the application of a regulatory strategy—namely, the direct-access answers—on our understanding of the strategic regulation of accuracy in memory reporting.

## The regulation of accuracy

The regulation of accuracy is a process that allows accuracy to be maximized. In general, it involves two basic processes: monitoring the answer, and a control process to decide what to do with it (e.g., Koriat & Goldsmith, 1996). Several models have elaborated on this basic idea. For example, the satisficing

---

✉ Karlos Luna  
karlos.luna.ortega@gmail.com

<sup>1</sup> Department of Psychology, Süleyman Şah University, Fatih Mah. Eski Ankara Asfaltı Üzeri 28, 34956 Orhanlı / Tuzla, Istanbul, Turkey

<sup>2</sup> Faculty of Psychology, University of the Basque Country, Gipuzkoa, Spain

<sup>3</sup> School of Psychology, Flinders University, Adelaide, South Australia

model of the regulation of accuracy (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; see also Koriat & Goldsmith, 1996) states that participants first rate the subjective likelihood that the alternatives (either presented, as in a recognition test, or self-generated, as in a recall test) are correct (monitoring process). Then, in the control process the likelihood of the best alternative is compared against a preestablished criterion (a confidence criterion). If it reaches the criterion, the answer is reported. If not, regulatory strategies are applied to maximize accuracy.<sup>1</sup>

More recently, Ackerman and Goldsmith (2008) proposed the dual-criterion model, which added an informativeness criterion to the control process. The informativeness criterion requires that the answer must transmit a minimum amount of information to be reported. For example, the answer that a robber is between 1 and 3 m tall is, although correct, unattractive because it conveys little information, and thus transgresses social and communicative norms (Grice, 1975). Again, when the best alternative does not meet the confidence or the informativeness criteria, or both, regulatory strategies are applied.

When the control process reaches a negative decision about the suitability of the answer, regulatory strategies are applied. There are three main regulatory strategies. One of these strategies is permitted by the report option, which gives participants the option to report or withhold their answer (Koriat & Goldsmith, 1996). By withholding answers that are subjectively evaluated as having a low probability of being correct, participants can increase their global accuracy.<sup>2</sup> Research into the use of this strategy has been conducted in educational settings, with test-takers being given the chance to either answer or not answer a question (Arnold, Higham, & Martín-Luengo, 2013; Higham, 2007), in the study of gambling behavior (Lueddeke & Higham, 2011), and in the eyewitness memory context (Evans & Fisher, 2011; Higham, Luna, & Bloomfield, 2011). Also, theoretically focused research designed to explore the contributions of different metacognitive components has used the report option (Arnold, 2013).

Another strategy that participants can apply is to vary the grain size of the answer, with participants being able to decide on the level of detail or precision of the answer that they provide (Goldsmith et al., 2002). For example, one participant may answer correctly that a robber was between 1.70 and

1.80 m tall, but another one may also be correct by answering that the robber was between 1.60 and 1.85 m tall. Effective regulation of the grain size of a memory report also increases accuracy (Ackerman & Goldsmith, 2008; Pansky & Nemets, 2012; Yaniv & Foster, 1995, 1997). Indeed, when mock-witnesses to a crime were given the option to use the grain-size strategy, the global accuracy of their reports increased (Goldsmith, Koriat, & Pansky, 2005; Weber & Brewer, 2008).

A third strategy that participants can apply to regulate the accuracy of an answer is to vary the plurality of the response (Luna, Higham, & Martín-Luengo, 2011). The plurality option involves the selection of a different number of alternatives, in much the same way that participants select the level of precision with the grain-size option. For example, for the question “How did the robber conceal her face? with a mask, a scarf, a stocking, a bandanna, or a balaclava,” a participant can select one alternative (mask; single answer), or three (mask, balaclava, and bandanna; plural answer), depending on her knowledge, to adjust the likelihood that the answer will include the correct alternative (mask). In both classroom (Higham, 2013) and eyewitness (Luna et al., 2011; Luna & Martín-Luengo, 2012b) contexts, research on the plurality option has shown that accuracy increases when participants are allowed to decide on the number of alternatives in the answer.

### Direct-access answers

In the original account of the dual-criterion model, which focused on the grain-size regulatory strategy, the authors stated that, for some questions, an answer can be retrieved at a “very precise level and with high confidence.” In these cases, they assumed that “the answer will simply be provided ‘as is’, with no grain adjustment needed” (Ackerman & Goldsmith, 2008, p. 1227, note 3). Ackerman and Goldsmith did not further discuss how those answers might affect the regulation of accuracy. Here we examined this issue further, focusing on the answers that do not need to go through a regulatory strategy. How do they affect the benefits associated with the regulation of accuracy?

Our main idea is that a subset of answers may be directly accessed at a level that makes the activation of regulatory strategies unnecessary. But, what is the nature of these direct-access answers? We propose that a direct-access answer is one that should be (a) retrieved rapidly and fluently, (b) recovered automatically (i.e., without elaborate processes such as discarding options known to be false), and (c) held with very high confidence. For example, when asked for the name of Earth’s satellite, some people may retrieve “the Moon” rapidly and may rate this alternative with very high confidence. In this case there is no need to apply any regulatory strategy, and the answer is reported as is. For this reason, when theoretical or applied aspects of the regulation of accuracy are studied, it is relevant to distinguish direct-access

<sup>1</sup> We draw a distinction here between the regulation of accuracy as a general process and the specific regulatory strategies. The regulation of accuracy encompasses the full sequence that includes retrieval, monitoring, control, and the selection of the answer. The specific regulatory strategies (i.e., report, grain size, and plurality option, explained below) are part of the control process and are activated only in certain circumstances (see the next section).

<sup>2</sup> *Confidence ratings* and *probability ratings* refer to different theoretical concepts. However, Luna, Higham, and Martín-Luengo (2011) showed that participants do not distinguish between them in the plurality option context. For this reason, we use both terms interchangeably here.

answers that do not require the application of a regulatory strategy from those that do require it.

Direct-access answers, as defined here, are likely the outcome of a retrieval process that is qualitatively different from that of other answers that go through a regulatory strategy. Several different retrieval processes may produce a direct-access answer: for example, a recollection process in an episodic task, or the use of decision-making System 1 in a semantic task. System 1 is automatic, effortless, fast, and intuitive (Kahneman, 2003) and is likely to produce answers that are reported rapidly and confidently (Kahneman, 2011). Retrieval processes are relevant because they can be used to control memory performance (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Shimizu & Jacoby, 2005) and may affect posterior processes (e.g., Butterfield & Metcalfe, 2006). Halamish, Goldsmith, and Jacoby (2012) pointed out that both retrieval (front-end) and postretrieval (back-end) processes may be used to control memory reporting. In this research, we focused on back-end processes, with the specific front-end processes that produce a direct-access answer being of secondary importance. What was important here was that direct-access answers are likely the product of a different retrieval process and would not need to go through a regulatory strategy.

Directly accessing an answer does not mean that the correct answer is always accessed (see the consensuality principle: Koriat, 2008, 2012). A very familiar or highly available incorrect answer could be accessed directly and reported with high confidence. For example, in the question “What is the capital city of Australia,” most people will answer Sydney, because it is the best-known Australian city, instead of Canberra, the correct answer. Despite the occasional direct access to an incorrect answer, high accuracy is generally expected for questions with a direct-access answer.

In addition, distinguishing between direct-access answers and those requiring the application of a regulatory strategy has applied relevance. For example, in the eyewitness memory and reporting setting, if a unit of information in a witness’s testimony could be identified as a direct-access answer and, hence, highly likely to be accurate, it might be given priority in the investigation over other units that are the outcome of a regulatory strategy. It might also be useful in educational contexts. If a student can identify enough direct-access answers to pass an exam, she would not need to answer more questions and would avoid the risk of a penalty for any incorrect answers.

Despite the theoretical acknowledgement of the existence of the direct-access answers and their potential applied relevance, no research examining the strategic regulation of memory reports has tried to disentangle direct-access answers from answers that are the product of a regulatory strategy. Without disentangling these different types of answers, it is unclear to what extent direct-access answers could have been driving what has been perceived as the outcome of effective regulatory strategies. Keeping the plurality-option terminology, a

strategy is effective for accuracy regulation if single answers are reported when their perceived likelihood of being correct is high, and rejected when that perceived likelihood is low (Luna et al., 2011). The larger the difference in accuracy between single selected and single rejected answers, the more effective the strategy and the better the regulation of accuracy should be. Because direct-access answers are expected to have high accuracy accompanied by high subjective probability of being correct, most, if not all, of the direct-access answers should be selected at a single level, and the remaining answers at a plural level. Thus, the decision to report or reject the single answer could have been more about distinguishing the direct-access answers from the rest, without the need to invoke a regulatory strategy to explain the increase in accuracy. In support of this idea, when an answer is rated with the highest confidence, as with a direct-access answer, resolution (i.e., the ability to distinguish correct from incorrect answers) is also high (Higham, Perfect, & Bruno, 2009; Payne, Jacoby, & Lambert, 2004). If regulatory strategies are not needed to explain the increase in accuracy, then previous studies about the regulation of accuracy may have not studied the effectiveness, usefulness, and benefits of regulatory strategies, but rather the ability to distinguish direct-access answers from other answers.

Even though we doubt that all of the single answers in previous studies were direct-access answers, the point is that there is no way to know the extent to which past results were affected by direct-access answers. One possibility is that the accuracy increase in previous studies was caused solely by the distinction between direct-access and other answers. If this is true, there would be no necessity to invoke the application of regulatory strategies. In that case, current theoretical explanations based on research with those strategies should be revised or discarded, and new theories should be proposed about how and how well direct-access answers are identified. Applied research should also change its focus to the direct-access answers and how they can be used to increase accuracy.

Another, more likely possibility is that the increase of accuracy in previous studies was caused by the combined operation of regulatory strategies and direct-access answers. If this is true, then two effects should be disentangled and studied separately. In this research, we examined this second alternative and tried to separate the contributions of the direct-access answers and the regulatory strategies.

### The present research

The present research had two main objectives. The first was to identify a subset of answers that are the product of a different retrieval process for which direct access is likely. The second was to test whether regulatory strategies increase accuracy even when the contribution of the direct-access answers is removed. To accomplish those objectives, we designed an

experiment using the plurality option and two reporting phases: a free-recognition and a forced-recognition phase. In the *free-recognition phase*, participants were given the chance to select as many alternatives as they wanted, instead of being artificially constrained by the experimenter to a certain number (as is usually the case). No research with the plurality option and a free-recognition phase has been reported before. Then, participants completed the standard plurality-option procedure (*forced-recognition phase*), with a few modifications.

## Experiment 1

### Method

**Participants** To determine the sample size, we first examined the effect sizes of the analyses that had shown regulation of accuracy in previous studies with the plurality option. Luna et al. (2011) reported Cohen's  $d$ s = 1.37 and 0.70 for their Experiments 1 and 2, respectively. Power analysis suggested a minimum sample size of 18 participants to find an effect size of Cohen's  $d = 0.70$ , with  $\alpha = .05$  and  $1 - \beta = .80$ . Twenty-six students (21 females, five males; mean age 19 years,  $SD = 1.23$ , range 18–22) from the University of Minho (Portugal) took part in this experiment in exchange for course credits. The participants were randomly allocated to one of the four counterbalanced conditions (see below).

**Materials** Forty general-knowledge questions were selected from a larger database available on the Internet.<sup>3</sup> The original questions had four alternatives, and two more were added by the authors of the present study. The questions addressed different domains, such as cinema, mythology, sports, geography, or history, to name a few. See the Appendix for a sample question. Two questionnaires with the questions in different orders were created for counterbalancing purposes.

**Procedure** Participants were tested individually or in small groups of up to four. They entered the lab and sat in front of a computer that first presented basic demographic questions, then the instructions, and finally the experimental phase with the questions (see the example in the Appendix).

The instructions framed the experiment in the context of a Q&A game like Trivial Pursuit with a friend. The instructions mentioned that 40 general knowledge questions with six alternative answers were to be presented, that only one of the alternatives was correct, and that their task was to select the

correct alternative. The instructions also mentioned several sections. In the first section, the free-recognition phase, participants had to select between one and five alternatives and rate the probability that the correct answer was one of those selected, on a 0 %–100 % scale. In the second section, the forced-recognition phase, they had to select (a) two alternatives and rate the probability that they had selected the correct answer (small answer), and (b) four alternatives and also rate the probability that they had selected the correct answer (large answer). In the original plurality-option procedure, single answers included only one alternative, and plural answers three alternatives. Because a two-alternative answer cannot be considered “single,” we changed the terminology here to *small* and *large* answers. A small answer was defined as having two alternatives in order to allow for a better exploration of the regulation of accuracy with and without the effect of the direct-access answers. If small answers in the forced-recognition phase had been defined as having one alternative, after removing the direct-access answers, only a few answers might have remained, which could make the analyses nonviable. To maintain the two-alternative difference between small and large, the latter condition was defined here as having four alternatives. The last task in the forced-recognition phase was to indicate whether participants preferred to answer with the two-alternative (small) or the four-alternative (large) answer to a friend who had asked that question.

Below the instructions, a sample question was presented with the same display from the experimental phase. The example included the answers of a potential participant and a brief explanation for her choices. The explanations also helped participants become familiar with the law of probability that the more alternatives were checked, the higher the probability of being correct, and vice versa (see the sentences in brackets in the Appendix).<sup>4</sup>

Participants read the instructions at their own pace. After that, the first experimental question was presented. The presentation of the questions, alternatives, and sections was the same as in the Appendix, but without alternatives selected and additional explanations. A bluish bar separated the free-recognition and forced-recognition phases. In the forced phase, for half of the participants Section A asked for two alternatives and Section B for four alternatives, and the opposite was true for the other half. Section A was always presented first. The order of presentation of the free and forced phases was not counterbalanced to avoid, under a forced-first, free-second order, participants in the free phase selecting the same number of alternatives as in the forced phase. Thus, the free-recognition phase was always presented first and the forced

<sup>3</sup> The original database can be found at [www.prof2000.pt/users/avcultura/cultGeral/Problema02Listagem.htm](http://www.prof2000.pt/users/avcultura/cultGeral/Problema02Listagem.htm) (in Portuguese). Used with kind permission of the site's administrator.

<sup>4</sup> This law holds for honest respondents who are trying to report the correct answer, as in the typical memory experiment in the laboratory. Someone deliberately choosing incorrect answers might produce a different pattern.

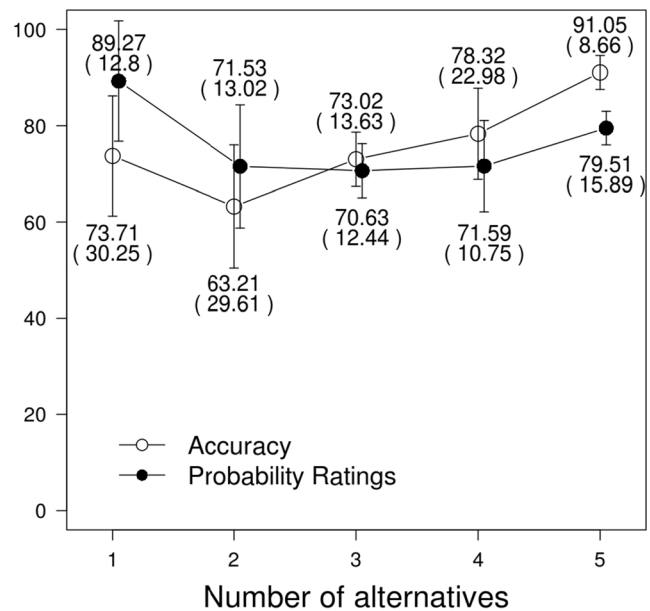
second. It was still possible that participants might first complete the forced phase (in the bottom part of the screen) and then the free phase (in the upper part), but this was highly unlikely, as to do so they would have had to scroll down the page and then up again.

## Results

First we present the results of both the accuracy and probability ratings for the free-recognition phase, followed by those of the forced-recognition phase. The objective of the free-recognition phase was to identify a subset of answers that were likely direct-access answers. The effect of the direct-access answers was controlled in the forced-recognition phase by removing the data points that corresponded to the direct-access answers in the free phase. For example, if in the free-recognition phase Participant 3 answered Question 2 with one alternative and confidence 100, then that answer was identified as a direct-access answer. To control for the effect of the direct-access answers in the forced-recognition phase, the answers of Participant 3 to Question 2 were removed.

Accuracy is presented in percentages to allow for direct comparison of the probability ratings. Unless otherwise stated,  $\alpha = .05$  and the reported confidence intervals (CIs) are 95 %. Cohen's  $d$  (henceforth,  $d$ ) was used to compute the effect sizes for pairwise comparisons; 95 % CIs around  $d$  are also reported.

**Free-recognition phase** This phase allowed the identification of a subset of answers that were the results of direct retrieval. Participants selected one, two, three, four, or five alternatives 9, 12, 25, 22 and 32 percent of the time ( $SDs = 7, 8, 11, 8,$  and  $19$ ), respectively. Accuracy and probability ratings as a function of the number of alternatives selected are displayed in Fig. 1. The distribution of responses suggests that the questions were difficult and that a direct-access answer was probably available for only a small portion of the questions and participants. We defined a direct-access answer as an answer with one alternative and the highest possible confidence—that is, rated with confidence 100. Other operational definitions of a direct-access answer are also possible, but these were discarded. For example, a more liberal definition would have been those answers selected with one alternative in the free-recognition phase. However, this definition included answers rated with a confidence less than 100, which thus were more likely to be the product of processes other than direct access. A stricter operational definition would have been that direct-access answers were those with one alternative, rated with confidence 100, and that were correct. However, this definition would miss the point that a direct-access answer does not always have to be correct. In any case, a replication of the analyses with different definitions did not change the main results and conclusions.



**Fig. 1** Mean accuracy and probability ratings in the free-recognition phase as a function of the number of alternatives selected in Experiment 1. Standard deviations are in parentheses, and error bars represent 95 % confidence intervals.  $Ns = 25, 23, 25, 25,$  and  $26$  for one to five alternatives

Only 5 % of the answers (60 % of the one-alternative answers) met our definition of a direct-access answer. As expected, accuracy for those answers was very high, although not at ceiling ( $M = 91.40, SD = 16.00, CI [83.69, 99.11], N = 19$  because some participants did not make direct-access answers).

To test whether accuracy for the free-recognition data was affected by the direct-access answers, we compared the accuracy for one-alternative answers rated with confidence 100—that is, direct-access answers—against the accuracy for one-alternative answers rated with confidence 90 or lower ( $M = 41.30, SD = 43.75, CI [17.06, 65.54], N = 15$  participants). Only nine participants had data in both conditions, making parametric statistical comparisons unreliable, but both Student's  $t$  test and its nonparametric counterpart the Wilcoxon test showed significant differences,  $t(8) = 2.86, p = .021, d = 1.57, CI [0.55, 2.54]$ , and  $V = 33, p = .040$  (for this analysis,  $M = 87.78, SD = 19.86, CI [72.51, 103.05]$ , and  $M = 34.44, SD = 43.91, CI [0.68, 68.20], N = 9$ ). More importantly, the confidence intervals for both scores did not overlap, strongly suggesting that direct-access answers, as defined here, were different from other answers and that, when pooled together, their accuracy was increased.

A similar comparison with probability ratings would not be of use, because they were part of the definition used to identify direct-access answers. As an indirect way to test the effect of direct-access answers on probability ratings, we compared the ratings for all of the answers selected with one alternative against the ratings for all of the answers selected with two

alternatives. This analysis was based on the basic probability law that the chances of selecting the correct alternative should be higher when more alternatives were selected. Thus, the ratings displayed in Fig. 1 should display a constant increase from selecting one alternative (base-rate chance of 16 percent) to five alternatives (base-rate chance of 84 %). However, Fig. 1 shows a peak in probability ratings with one alternative (and also that participants ignored the effect of the base-rate probability, in line with previous research; see Higham, 2013). Consistent with the idea that direct access to the answer increases the probability ratings of all of the one-alternative answers, the perceived probability of being correct was higher with one than with two alternatives,  $t(22) = 5.57, p = .016, d = 1.29, CI [0.73, 1.85]$ .

An alternative way to test the existence of a separate set of one-alternative answers that are the product of a different retrieval process is by using Type-2 receiver operating characteristic (ROC) curves. A Type-2 ROC curve plots correct and incorrect responses for different confidence levels (Higham, 2007; Macmillan & Creelman, 2005). Figure 2 shows the rates of correct and incorrect responses for all of the one-alternative answers with confidence 0 or higher (0+), 10 or higher (10+), and so on. If one-alternative answers came from one, single, normal distribution, the curve should intercept the y-axis of correct responses approximately at or close to the origin. If there are two distributions, the curve should intercept the y-axis above the origin, with the interception point estimating the proportion of items in the second distribution. In a Type-2 ROC curve, it is not usual that the most conservative point (confidence 100) has 0 incorrect responses. However, it is possible to visually extrapolate the shape of the curve to see where it intercepts the y-axis. As can be seen, the extrapolation of the Type-2 ROC curve does indeed intercept the y-axis. In addition, the low number of incorrect responses at the most

conservative point also suggests that there are two separate distributions, one with high accuracy (high rate of correct responses and low rate of incorrect responses) coupled with high confidence, and another with lower accuracy (high rates of both correct and incorrect responses) coupled with low confidence. Both distributions likely overlap with direct-access answers and other answers, but the matching is probably not perfect, because the y-axis of the Type-2 ROC curve shows the proportion of correct responses, but sometimes a direct-access answer may be incorrect.

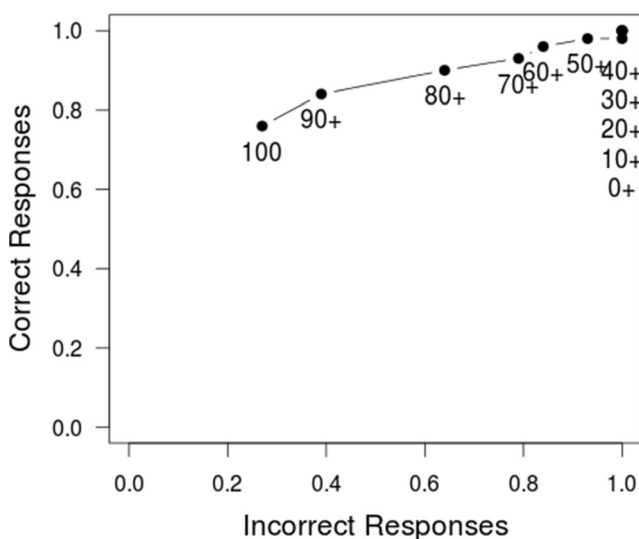
In general, these results suggest that both accuracy and the probability ratings of one-alternative answers in the free-recognition phase were affected by a small subset of direct-access answers. Now, we turn our attention to the effects of the direct-access answers on the regulation of accuracy in the standard plurality-option paradigm (forced-recognition phase).

**Forced-recognition phase** Two sets of analyses were conducted to examine the regulation of accuracy with and without direct answers. The main statistics for all data are reported in Table 1.

When all of the answers were examined, participants selected the small and large answers 33 % and 67 % ( $SD = 16$ ) of the time, respectively. The regulation of accuracy with the plurality option is shown when a participant reports a small answer when its accuracy is high but rejects it when its accuracy is low. Accuracy was higher for small-selected than for small-rejected answers,  $t(25) = 4.49, p < .001, d = 1.35, CI [0.81, 1.88]$ . As a consequence of the strategic regulation of the number of alternatives and the occasional selection of the large answer, accuracy increased from what could have been obtained if only small answers were allowed (48.17) to that finally obtained (72.98),  $t(25) = 14.96, p < .001, d = 3.02, CI [2.10, 3.93]$ .

Then, the effect of the direct-access answers on the regulation of accuracy was removed by eliminating the answers in the forced phase corresponding to the 5 % of one-alternative answers with confidence 100 in the free phase. These main statistics are reported in Table 2. One participant was lost for the following analyses because, after this elimination, she had no small-selected answers. In this data set, participants selected the small and the large answers 31 % and 69 % ( $SD = 17$ ) of the time, respectively. As expected, participants selected more small answers with direct-access answers than without,  $t(24) = 5.20, p < .001, d = 0.20, CI [-0.19, 0.60]$ . The rest of the analyses replicated those with all of the data. The accuracy was higher for selected small answers than for rejected small answers,  $t(24) = 2.86, p = .009, d = 0.88, CI [0.42, 1.35]$ , and accuracy increased from all small (46.23) to the finally selected answer (71.54),  $t(24) = 15.54, p < .001, d = 3.23, CI [2.23, 4.21]$ . In summary, after the removal of the direct-access answers, the results show that participants were able to regulate the number of alternatives and increased their final accuracy.

Finally, a comparison between the ability to regulate accuracy with and without a direct-access answer was conducted.



**Fig. 2** Type-2 ROC curve for Experiment 1: Data from answers with one alternative in the free-recognition phase

**Table 1** Means, with (standard deviations) and [95 % confidence intervals], for accuracy and probability ratings in the forced-recognition phase of Experiment 1 for all data

	Selected Answers	Rejected Answers	Total	All Selected
Accuracy				
Small answer (two alternatives)	62.23 (17.58) [55.12, 69.34]	43.07 (9.73) [39.14, 47.00]	48.17 (7.16) [45.29, 51.05]	72.98 (9.14) [69.29, 76.67]
Large answer (four alternatives)	79.95 (8.08) [76.70, 83.20]	85.72 (12.63) [80.61, 90.83]	81.63 (6.85) [78.87, 84.39]	
Probability Ratings				
Small answer (two alternatives)	73.35 (11.79) [68.58, 78.11]	42.07 (11.29) [37.51, 46.63]	51.19 (10.96) [46.76, 55.62]	67.67 (11.50) [63.03, 72.32]
Large answer (four alternatives)	66.17 (12.83) [60.98, 71.35]	82.14 (10.63) [77.84, 86.43]	70.82 (44.42) [66.21, 75.44]	

$N = 26$

An analysis of variance (ANOVA) was not considered appropriate, because of the different rates of selection of small answers between the two data sets. The difference between the accuracy results for small selected and small rejected was computed for all data and without direct-access answers. This measure was higher, meaning that the regulation of accuracy was easier, for all data ( $M = 17.55$ ,  $SD = 20.57$ ,  $CI [9.07, 26.03]$ ) than when direct-access answers were removed ( $M = 11.90$ ,  $SD = 20.82$ ,  $CI [3.31, 20.49]$ ):  $t(24) = 4.25$ ,  $p < .001$ ,  $d = 0.27$ ,  $CI [-0.13, 0.67]$ .

## Discussion

Current theoretical models of the regulation of accuracy acknowledge the existence of direct-access answers that are fast, fluent, automatically retrieved, and held with high confidence. However, this is the first time that their role in the regulation of accuracy has been tested. In Experiment 1, we identified a subset of answers that were likely to be the consequences of direct access to the answer. As predicted, accuracy and probability ratings for the direct-access answers were very high. In our sample, only 5 % of all of the answers were identified as

direct-access answers, but they had a significant effect on the regulation of accuracy. In particular, our results showed that the direct-access answers made the regulation of accuracy easier.

Our results also showed that the regulation of accuracy still increased accuracy when direct-access answers were removed. This result suggests that regulatory strategies also contribute to increased accuracy and that the results of previous studies cannot merely be explained by a discrimination between direct-access and non-direct-access answers. Regulatory strategies are useful and have actual positive consequences for memory reports, but the effect of direct-access answers should be taken into account when the memory benefits of the regulation of accuracy are considered.

An alternative explanation that could account for some of the results of the forced-recognition phase is that one-alternative answers are distributed normally—that is, that there is not a separate group of direct-access answers—and that direct-access answers are only the upper tail of this distribution. If the upper tail is removed, as when we removed the direct-access answers, then the mean of the distribution should go down, and resolution (i.e., the ability to distinguish correct

**Table 2** Means, with (standard deviations) and [95 % confidence intervals], for accuracy and probability ratings in the forced-recognition phase of Experiment 1 when direct-access answers were removed

	Selected Answers	Rejected Answers	Total	All Selected
Accuracy				
Small answer (two alternatives)	55.07 (16.16) [48.40, 61.74]	43.17 (9.91) [39.08, 47.26]	46.23 (6.97) [43.36, 49.10]	71.54 (8.62) [68.00, 75.10]
Large answer (four alternatives)	79.69 (8.13) [76.33, 83.05]	82.79 (15.02) [76.60, 88.98]	80.60 (6.96) [77.73, 83.47]	
Probability Ratings				
Small answer (two alternatives)	67.72 (11.70) [62.88, 72.55]	41.44 (11.05) [36.88, 46.00]	48.55 (11.54) [43.79, 53.32]	66.12 (11.90) [61.21, 71.03]
Large answer (four alternatives)	65.98 (13.06) [60.59, 71.37]	78.71 (11.59) [73.93, 83.49]	69.57 (12.04) [64.60, 74.54]	

$N = 25$

from incorrect answers) should be lower. This is exactly what we found when direct-access answers were removed. However, the Type-2 ROC curve suggests that there was more than one distribution. Therefore, the decrease in resolution was not caused by removing the upper tail of one distribution, but rather by removing part or the whole of a different distribution.

One limitation of this experiment is that we cannot be completely sure that our operational definition identified only direct-access answers. Other postretrieval processes could have led to one-alternative answers with 100 % confidence that were not a direct-access answers. For example, it could be that for a given question a participant was able to identify five alternatives as incorrect with high certainty and to select the remaining one with the highest confidence because it had to be correct. Our data do not allow us to completely rule out this possibility, but we doubt that this was the case for a significant portion of the direct-access answers identified here. The questions were selected from a large pool so that all of the alternatives seemed similar, and it would have been difficult to eliminate all of the incorrect alternatives without knowing the correct answer.

A potential solution to distinguish between an automatic process, such as direct retrieval, and a more elaborate process, such as the elimination of alternatives, could be provided by response times (RTs). Direct-access answers should be retrieved quickly, and the more elaborate processes should take longer. In Experiment 2, we replicated the procedure of Experiment 1, with several changes, and measured RTs.

## Experiment 2

Our theoretical definition of a direct-access answer states that it should be retrieved rapidly and should be held with high confidence. In Experiment 2, we measured RTs and incorporated them into our operational definition of a direct-access answer. We also changed the materials and questions to examine whether the effects of the direct-access answers found in Experiment 1 would also be found in an eyewitness memory scenario.

## Method

Several changes were made to the procedure used in Experiment 1. Instead of 40 general-knowledge questions, 32 questions about a video of a bank robbery were used, and instead of presenting six alternatives for each question in the test, only five were used. Finally, we recorded the RTs in the free phase.

**Participants** Thirty-three paid students (23 females, ten males; mean age = 21.27 years,  $SD = 2.78$ , range 18–29) from Flinders University took part in this experiment. Only the

order of the sections (two alternatives and four alternatives) was counterbalanced. Participants were randomly allocated to the counterbalanced conditions.

**Materials** A 3-min video from Luna and Migueles (2008, 2009) was used. The video was an excerpt from the film *The Stick-Up* and showed a bank robbery. In the video, two security guards took some sacks of money to the safe deposit room in a bank and drove away. A bank robber stationed nearby cut off the power supply to the building, walked into the bank in disguise, threatened the people inside and made off with the money. Thirty-two questions with five alternatives (one correct and four incorrect) were developed for this experiment. All of the alternatives were plausible, addressed forensically relevant details, and were either central to the scene (e.g., The robber used a mask to conceal his face. What kind of mask?) or peripheral to the main action (e.g., When the robber went away, who approached the door to have a look outside?).

**Procedure** Other than the differences mentioned, the basic procedure and the instructions were the same as in Experiment 1, with a few adaptations for the new context. For example, selection of all of the five alternatives was not allowed. The main change was related to how the different sections were displayed. To allow for the recording of RTs in the free-recognition phase, questions were presented on three different screens. A first screen presented the questions and the five alternatives and prompted participants to select as many as they wanted, from one to four (free-recognition phase). RTs were measured from when this screen was displayed to when participants clicked a button to go to the next screen. Then, a second screen displayed the question and the alternatives, with the selections of the participant underlined and in bold font. Below that, a confidence rating from 0 to 100 was requested (confidence of the free-recognition phase). Finally, a third screen presented the forced-recognition phase in the same way as in Experiment 1. Seventeen participants completed, in the forced-recognition phase, the two-alternative section first and then the four-alternative section. The other 16 completed these phases in the opposite order.

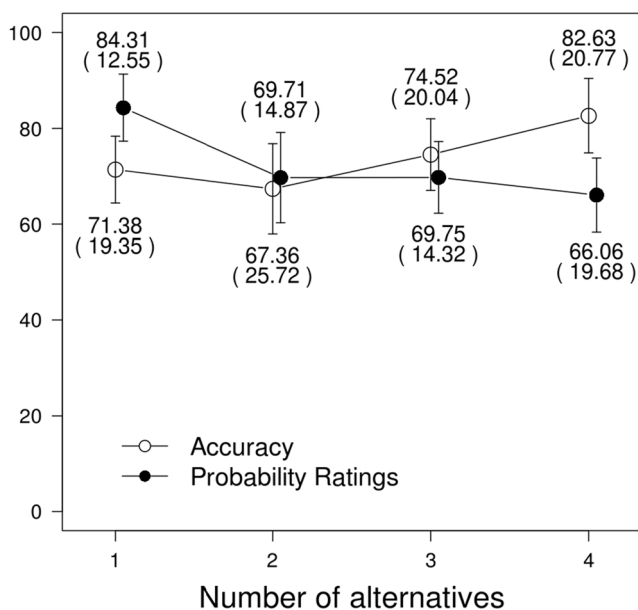
## Results

We present first the results of the free recognition, including RTs, and, second, those of the forced-recognition phase. In the forced phase, three sets of analyses were conducted, the first with the full sample, the second with the same operational definition of direct access as in Experiment 1 (answers with one alternative and confidence 100), and the third with a more restrictive definition (the faster half of the answers with one alternative and confidence 100). The results replicated those of the Experiment 1.



**Free-recognition phase** Participants selected one, two, three, or four alternatives 41 %, 18 %, 19 %, and 22 % of the time ( $SDs = 23, 11, 11, \text{ and } 22$ , respectively). Accuracy and probability ratings as functions of the number of alternatives selected are displayed in Fig. 3. The distribution of responses suggests that the questions were, in general, easier than those in Experiment 1 and that more answers could be the product of a direct access. To examine the RTs, outliers were identified for each participant that were three  $SDs$  higher or lower than the participant's mean RT. Only 12 RTs (1.1 % of the 1,056 answers) were identified as outliers and removed from all of the analyses. The RTs for answers with one to four alternatives was 10.97, 15.80, 17.38, and 18.56 s ( $SDs = 3.11, 4.59, 5.07, \text{ and } 6.03$ , respectively). Answers took more time when more alternatives were selected because participants had to click more times to provide their answer.

To examine the relationship between RT and confidence, we focused on the answers with one alternative because they were the only ones that could follow direct access. In support of the idea that direct-access answers should be faster and rated with high confidence, the correlation between confidence and RT was negative and significant,  $r = -.26, p = .002$ . In addition, RTs were faster for answers with one alternative and rated with confidence 100 ( $M = 9.16, SD = 3.23, CI [7.76, 10.58]$ ) than for one alternative rated with confidence 90 ( $M = 10.73, SD = 4.23, CI [8.98, 12.47]$ ), both  $Ns = 23$ ),  $t(22) = 2.47, p = .022, d = 0.42, CI [-0.01, 0.83]$ , or confidence 80 ( $M = 12.12, SD = 4.80, CI [10.00, 14.26]$ ,  $N = 22$ ),  $t(21) = 4.12, p < .001, d = 0.68, CI [0.21, 1.14]$ . Comparisons with other confidence levels had  $Ns < 15$ .



**Fig. 3** Mean accuracy and probability ratings in the free-recognition phase as a function of the number of alternatives selected in Experiment 2. Standard deviations are in parentheses, and error bars represent 95 % confidence intervals.  $Ns = 32, 31, 30, \text{ and } 30$  for one to four alternatives

We operationalized direct access in two different ways. First, as in Experiment 1, we computed answers with one alternative and confidence 100. Twenty percent of all of the answers met these prerequisites (47 % of the answers with one alternative). Accuracy was higher for those answers than for answers with one alternative and confidence 90 or less ( $M = 91.07, SD = 12.05, CI [86.48, 95.66]$ , and  $M = 50.52, SD = 27.84, CI [39.93, 61.11]$ , respectively,  $N = 29$ ),  $t(28) = 8.16, p < .001, d = 1.89, CI [1.27, 2.50]$ .

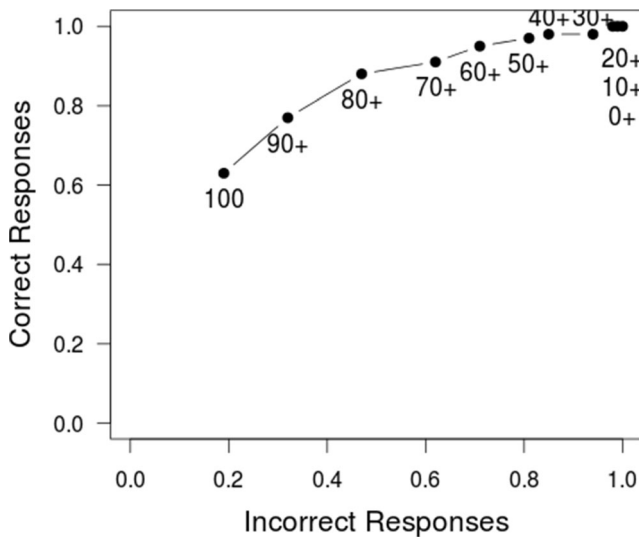
The second operationalization included the RT. The median RT for answers with one alternative and confidence 100 was computed for each participant, and the answers faster than or equal to the median were identified. The faster half of the answers with one alternative and confidence 100 were used because they identified a proportion of answers that allowed conducting the analyses. More conservative operationalizations (e.g., the fastest 25 % of answers) identified too small a subset of answers, which, in turn, made analyses nonviable. Ten percent of all the answers (25 % of the one-alternative answers) met these prerequisites. Accuracy for those answers was also higher than for the rest of the answers with one alternative, including those rated with confidence 100 and RTs slower than the median ( $M = 92.83, SD = 12.81, CI [88.13, 97.53]$  and  $M = 63.81, SD = 25.59, CI [54.42, 73.20]$ , respectively, both  $Ns = 31$ ),  $t(30) = 6.90, p < .001, d = 1.43, CI [0.92, 1.93]$ .

Finally, also replicating the results of Experiment 1, probability ratings were higher for all of the questions selected with one alternative than for those selected with two alternatives,  $t(30) = 6.83, p < .001, d = 1.09, CI [0.64, 1.53]$ . In general, the results support the conclusion that direct-access answers have strong effects on both accuracy and probability ratings.

A Type-2 ROC curve was also computed with all of the one-alternative answers to examine the likely distribution of responses (see Fig. 4). The ROC curve also suggests that a subgroup of answers have high accuracy (high rate of correct responses and low rate of incorrect responses) and high confidence, thus suggesting two different response distributions.

**Forced-recognition phase** When all of the answers were analyzed, participants selected the small answers 59 % of the time ( $SD = 21$ ). The main statistics are reported in Table 3. Accuracy was higher for small-selected than for small-rejected answers,  $t(32) = 8.41, p < .001, d = 1.98, CI [1.38, 2.56]$ , and also higher for all of the selected answers than for all of the small answers,  $t(32) = 7.99, p < .001, d = 1.64, CI [1.11, 2.16]$ . These results show both the successful regulation of accuracy and the benefits for the final report.

When the answers with one alternative and rated with confidence 100 were removed from the sample, participants selected both small and large answers 50 % of the time ( $SD = 23$ ). The main statistics are reported in Table 4. The rate of selection of small answers was lower here than when all of the sample was analyzed,  $t(32) = 8.66, p < .001, d = 0.40, CI [0.03, 0.79]$ .



**Fig. 4** Type-2 ROC curve for Experiment 2: Data from answers with one alternative in the free-recognition phase

Accuracy was higher for small-selected than for small-rejected answers,  $t(32) = 5.51, p < .001, d = 1.32, CI [0.84, 1.78]$ , meaning that the regulation of accuracy was also successful. However, the difference between small-selected and small-rejected answers was lower here than for all of the sample,  $t(32) = 4.26, p < .001, d = 0.31, CI [-0.05, 0.65]$ , meaning that the regulation of accuracy was more difficult. Accuracy was also higher here for all of the selected answers than for all of the small answers,  $t(32) = 7.86, p < .001, d = 1.79, CI [1.23, 2.34]$ .

When the faster half of the one-alternative answers rated with confidence 100 were removed from the sample, participants selected the small answers 54 % of the time ( $SD = 22$ ). The main statistics are reported in Table 5. The results replicated those above. The rate of selection of small answers was lower here than when all of the sample was analyzed,  $t(32) = 10.16, p < .001, d = 0.19, CI [-0.16, 0.53]$ . Accuracy was higher for small-selected than for small-rejected answers,  $t(32) = 7.06, p < .001, d = 1.67, CI [1.13, 2.20]$ . The difference between small-selected and small-rejected answers was lower

here than for all of the sample,  $t(32) = 5.56, p < .001, d = 0.17, CI [-0.17, 0.51]$ , meaning that the regulation of accuracy was more difficult. Accuracy was also higher for all of the selected answers than for all of the small answers,  $t(32) = 8.44, p < .001, d = 1.71, CI [1.17, 2.25]$ .

**Discussion**

Despite changes in the materials and procedure, and despite a stricter operational definition of a direct-access answer, we replicated the main findings of Experiment 1. More importantly, the results of Experiment 2 showed that answers with one alternative and confidence 100 were faster than other answers with one alternative. This result suggests that it is unlikely that direct-access answers, as defined here, were the outcome of an elaborate postretrieval process such as discounting alternatives.

**General Discussion**

In two experiments, we tested the hypothesis that the increase in accuracy found in previous studies may have reflected performance on items for which participants had direct access to the answer, without the need to invoke a regulatory strategy. Several novel findings from this research can be highlighted. First, we found an effect of direct-access retrieval in both the accuracy and probability ratings of one-alternative answers. Second, the plurality option was useful for increasing accuracy even when the effect of the direct-access answers was removed. The results support the contribution of regulatory strategies, in this case the plurality option, to the increase in accuracy in the final report.

When an answer, either correct or incorrect, is directly accessed, participants tend to report it with high confidence without the need to activate a regulatory strategy. This research has been the first to identify this category of direct-access answers and to experimentally examine their effects

**Table 3** Means, with (standard deviations) and [95 % confidence intervals], for accuracy and probability ratings in the forced-recognition phase of Experiment 2 for all data

	Selected Answers	Rejected Answers	Total	All Selected
<b>Accuracy</b>				
Small answer (two alternatives)	78.24 (11.66) [74.11, 82.37]	46.47 (19.53) [39.54, 53.40]	64.02 (9.44) [60.68, 67.36]	79.55 (9.47) [76.19, 82.91]
Large answer (four alternatives)	83.57 (11.30) [79.56, 87.58]	94.94 (4.92) [93.19, 96.69]	90.25 (4.61) [88.62, 91.88]	
<b>Probability Ratings</b>				
Small answer (two alternatives)	84.79 (10.40) [81.10, 88.48]	51.36 (14.82) [46.11, 56.62]	69.51 (13.04) [64.88, 74.13]	77.51 (13.08) [72.87, 82.15]
Large answer (four alternatives)	68.87 (18.88) [62.18, 75.57]	85.99 (14.96) [80.69, 91.30]	78.24 (15.60) [72.71, 83.77]	

**Table 4** Means, with (standard deviations) and [95 % confidence intervals], for accuracy and probability ratings in the forced-recognition phase of Experiment 2 when answers with one alternative and confidence = 100 were removed

	Selected Answers	Rejected Answers	Total	All Selected
<b>Accuracy</b>				
Small answer (two alternatives)	70.13 (15.56) [64.61, 75.65]	45.61 (21.24) [38.07, 53.15]	57.04 (9.84) [53.56, 60.52]	76.23 (11.57) [72.14, 80.32]
Large answer (four alternatives)	83.14 (12.26) [78.80, 87.48]	92.42 (8.64) [89.36, 95.48]	87.87 (6.35) [85.61, 90.13]	
<b>Probability Ratings</b>				
Small answer (two alternatives)	78.45 (11.77) [74.28, 82.63]	50.55 (14.50) [45.41, 55.69]	63.22 (12.39) [58.83, 67.61]	72.91 (13.47) [68.14, 77.69]
Large answer (four alternatives)	68.43 (18.70) [61.80, 75.06]	81.44 (17.06) [75.39, 87.49]	74.47 (15.80) [68.87, 80.07]	

in the context of the regulation of accuracy. In general, the results showed that a subset of direct-access answers has strong effects on accuracy and probability ratings. The ability to identify a direct-access answer has potential consequences for the application of the regulation of accuracy. In situations in which there is uncertainty about the outcome, as when gambling or during a test, the ability to identify a subset of answers with very high accuracy with a simple rule such as those used here would be of practical relevance.

For example, consider the case of an exam with 30 multiple-choice questions, with one point for a correct answer and a penalty for error, and 15 points required to pass. The ability to know how many direct-access answers you reported may be of use here. A student might try a first pass through all of the questions, answering only those for which she was sure that she had identified the correct alternative. One of the ways of doing so would be to apply some of the criteria used here to identify direct-access answers: fast retrieval of a memory with high confidence. If 15 direct-access answers were identified and reported in the first round, the test taker could stop answering and avoid risking a penalty, because she would pass. In an exam, the regulation of the number of alternatives might not

be an option (but see Higham, 2013), but it could be in an eyewitness memory scenario with the appropriate instructions. If answering with two or more alternatives were explicitly allowed, a witness might sometimes select one alternative with high confidence and sometimes select two alternatives. Then, police could focus on those answers that were likely the consequences of direct access, because their accuracy should be very high. If the plurality option was not explicitly allowed, then the identification of direct-access answers might be more difficult, because all of the answers would include one alternative, and only the criteria of confidence and RT would remain. Although in some circumstances the relationships between confidence and accuracy (Luna & Martín-Luengo, 2012a) and between RT and accuracy (Weber & Brewer, 2003, 2006) are high, these might not be strong enough to identify a direct-access answer. Future research should test the conditions that could allow the identification of such answers in recognition tests when the regulation of accuracy with the report option, the grain size, or the plurality option was not allowed.

At a theoretical level, the dual-criterion model assumes that, when the informativeness of the answer increases, the probability that it is correct decreases (Ackerman &

**Table 5** Means, with (standard deviations) and [95 % confidence intervals], for accuracy and probability ratings in the forced-recognition phase of Experiment 2 when the fastest 50 % of answers with one alternative and confidence = 100 were removed

	Selected Answers	Rejected Answers	Total	All Selected
<b>Accuracy</b>				
Small answer (two alternatives)	74.64 (12.83) [70.10, 79.18]	46.71 (19.87) [39.66, 53.76]	60.79 (9.49) [57.43, 64.15]	77.77 (10.31) [74.12, 81.42]
Large answer (four alternatives)	83.77 (10.92) [79.90, 87.64]	94.15 (6.12) [91.17, 96.33]	89.29 (5.24) [87.44, 91.14]	
<b>Probability Ratings</b>				
Small answer (two alternatives)	82.51 (10.37) [78.83, 86.19]	51.25 (14.71) [46.03, 56.47]	66.76 (12.54) [62.13, 70.20]	75.54 (12.96) [70.95, 80.14]
Large answer (four alternatives)	68.79 (18.86) [62.10, 75.48]	84.63 (15.30) [79.20, 90.05]	76.56 (15.51) [71.06, 82.06]	

Goldsmith, 2008). This is true for objective probabilities, but our results suggest that this does not hold for objective accuracy or subjective probabilities. Some direct-access answers with high informativeness also have high objective accuracy and are rated with high subjective probabilities of being correct. However, this does not mean that the dual-criterion model should be amended. On the contrary, it highlights the need for controlling the direct-access answers. The most recent model of the regulation of accuracy acknowledges that, when an answer is directly accessed, participants do not need to engage in any regulatory strategy (Ackerman & Goldsmith, 2008, p. 1227, note 3). Despite this, to date, direct-access answers have not been controlled and have only been analyzed along with other answers that require the application of regulatory strategies. Our research points out the importance of distinguishing between them in order to examine the actual benefits and costs of the regulatory strategies in the regulation of accuracy.

The second novel finding is that accuracy increased even when the direct-access answers were removed from the analyses. An implication of the direct access to the answer is that this could have made regulatory strategies unnecessary. This was not, however, the case, although direct-access answers did make the regulation of accuracy easier. The difference between accuracy for small-selected answers, which included the direct-access answers, and small-rejected answers (i.e., the measure of the ability to regulate accuracy) was higher with direct-access answers than without. Despite this, our results also showed that the regulatory strategy of the plurality option increased accuracy even in the absence of direct-access answers. This shows that regulatory strategies are needed to explain the increase in accuracy. Our results also support the idea that the regulation of accuracy is a strong and reliable phenomenon, and also that it is replicable under different conditions (e.g., the different operational definitions of small and large in Luna et al., 2011, vs. here) and different materials (semantic memory in Exp. 1, episodic memory in Exp. 2). This flexibility actually increases the applicability of the plurality option, because the conditions, parameters, and prerequisites of, for example, the interrogation of a witness vary greatly. A technique that can improve accuracy in a wide set of circumstances could be of use in those cases.

Finally, our research highlights the importance of the definition of the number of alternatives of small and large answers. Sometimes a one-alternative answer could be of interest: for example, when a direct comparison is made with a test in which the regulation of accuracy is not allowed, or when a well-known memory phenomenon that involves standard one-alternative answers is being replicated in the context of a plurality-option study. However, at other times better experimental control of the regulation of accuracy might be needed. In this case, defining small answers as those with two alternatives might help to identify direct-access answers.

**Author note** This research was supported by an ARC Linkage International Social Sciences Collaboration grant and an ESRC (UK) Bilateral Australia grant, awarded to N.B., L. Hope, and F. Gabbert.

### Appendix

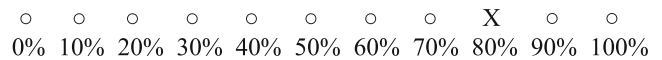
Example question presented in the instructions of Experiment 1. X denoted the selections of a potential participant. A brief explanation of her choices was included in brackets. During the experimental phase the same layout was used, without the selections and the text in brackets. In Experiment 2, the example question and the five alternatives were presented on three different screens mimicking those of the experimental phase, but the brief explanations in brackets in the example question were the same.

1.1. What was the name of the bomber that dropped the first atomic bomb over a Japanese city?

Select between 1 and 5 alternatives

- |   |   |
|---|---|
| <input type="checkbox"/> Black Widow          | <input checked="" type="checkbox"/> Free Star |
| <input type="checkbox"/> Bockscar             | <input checked="" type="checkbox"/> Enola Gay |
| <input checked="" type="checkbox"/> The Eagle | <input type="checkbox"/> Little Boy           |

Indicate the probability that you selected the correct alternative



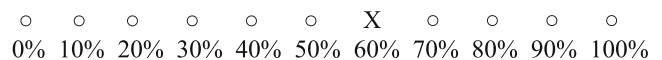
[The participant does not know which one is the correct alternative, but she can reject some alternatives that she knows are not correct. Thus, she selects three alternatives, with a probability of being correct of 80 percent.]

1.2. What was the name of the bomber that dropped the first atomic bomb over a Japanese city?

A. Select **two** alternatives that you think could be correct

- |                                      |   |
|--------------------------------------|---|
| <input type="checkbox"/> Black Widow | <input checked="" type="checkbox"/> Free Star |
| <input type="checkbox"/> Bockscar    | <input checked="" type="checkbox"/> Enola Gay |
| <input type="checkbox"/> The Eagle   | <input type="checkbox"/> Little Boy           |

Indicate the probability that you selected the correct alternative



[The participant has to select only two alternatives, so she selects the two with higher probabilities of being correct. The probability that the correct alternative included the correct answer decreased to 60 percent.]

B. Select **four** alternatives that you think could be correct

- |   |   |
|---|---|
| <input checked="" type="checkbox"/> Black Widow | <input checked="" type="checkbox"/> Free Star |
| <input type="checkbox"/> Bockscar               | <input checked="" type="checkbox"/> Enola Gay |
| <input checked="" type="checkbox"/> The Eagle   | <input type="checkbox"/> Little Boy           |

Indicate the probability that you selected the correct alternative

○ ○ ○ ○ ○ ○ ○ ○ ○ X ○  
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

[The participant has to select four alternatives, so she adds another one to the three that were selected in the first presentation of the question. The probability that the correct alternative included the correct answer increased to 90 percent.]

If a friend asked you this question, which answer (A -2 alternatives- or B -4 alternatives-) would you like to report?

X A ○ B

## 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.
- Arnold, M. M. (2013). Monitoring and meta-metacognition in the own-race bias. *Acta Psychologica*, *144*, 380–389.
- 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.
- Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*, 69–84.
- Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, *25*, 501–508.
- 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.
- 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
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York, NY: Academic Press.
- Halamish, V., Goldsmith, M., & Jacoby, L. (2012). Source-constrained recall: Front-end and back-end control of retrieval quality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1–15.
- 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.
- 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.
- Higham, P. A., Perfect, T. J., & Bruno, D. (2009). Investigating strength and frequency effects in recognition memory using type-2 signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 57–80.
- Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*, 852–857. doi:10.3758/BF03196776
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *9*, 697–720. doi:10.1037/0003-066X.58.9.697
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus, & Giroux.
- Koriat, A. (2008). Subjective confidence in one's answers: The consensuality principle. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 945–959.
- Koriat, A. (2012). When are two heads better than one and why? *Science*, *336*, 360–362.
- 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
- Lueddke, S. E., & Higham, P. A. (2011). Expertise and gambling: Using Type 2 signal detection theory to investigate differences between regular gamblers and nongamblers. *Quarterly Journal of Experimental Psychology*, *64*, 1850–1871.
- 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. (2012a). Confidence-accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology*, *26*, 289–295.
- 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.
- Luna, K., & Migueles, M. (2008). Typicality and misinformation: Two sources of distortion. *Psicológica*, *29*, 171–188.
- Luna, K., & Migueles, M. (2009). Acceptance and confidence of central and peripheral misinformation. *The Spanish Journal of Psychology*, *12*, 405–413.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory*. Mahwah, NJ: Erlbaum.
- Pansky, A., & Nemets, E. (2012). Enhancing the quantity and accuracy of eyewitness memory via initial memory testing. *Journal of Applied Research in Memory and Cognition*, *1*, 2–10.
- Payne, B. K., Jacoby, L. L., & Lambert, A. J. (2004). Memory monitoring and the control of stereotype distortion. *Journal of Experimental Social Psychology*, *40*, 52–64.
- Shimizu, Y., & Jacoby, L. L. (2005). Similarity-guided depth of retrieval: Constraints at the front end. *Canadian Journal of Experimental Psychology*, *59*, 17–21. doi:10.1037/h0087455
- Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology*, *88*, 490–499. doi:10.1037/0021-9010.88.3.490
- Weber, N., & Brewer, N. (2006). Positive versus negative face recognition decisions: Confidence, accuracy and response latency. *Applied Cognitive Psychology*, *20*, 17–31.
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, *14*, 50–60.
- Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, *124*, 424–432. doi:10.1037/0096-3445.124.4.424
- Yaniv, I., & Foster, D. P. (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making*, *10*, 21–32.