

The effects of divided attention at study and reporting procedure on regulation and monitoring for episodic recall



James Sauer^{a,*}, Lorraine Hope^b

^a University of Tasmania, Australia

^b University of Portsmouth, UK

ARTICLE INFO

Article history:

Received 25 August 2015

Received in revised form 6 April 2016

Accepted 27 May 2016

Available online xxxx

Keywords:

Recall

Memory

Metacognitive monitoring

Confidence

Eyewitness memory

Divided attention

ABSTRACT

Eyewitnesses regulate the level of detail (grain size) reported to balance competing demands for informativeness and accuracy. However, research to date has predominantly examined metacognitive monitoring for semantic memory tasks, and used relatively artificial phased reporting procedures. Further, although the established role of confidence in this regulation process may affect the confidence–accuracy relation for *volunteered* responses in predictable ways, previous investigations of the confidence–accuracy relation for eyewitness recall have largely overlooked the regulation of response granularity. Using a non-phased paradigm, Experiment 1 compared reporting and monitoring following optimal and sub-optimal (divided attention) encoding conditions. Participants showed evidence of sacrificing accuracy for informativeness, even when memory quality was relatively weak. Participants in the divided (cf. full) attention condition showed reduced accuracy for fine- but not coarse-grained responses. However, indices of discrimination and confidence diagnosticity showed no effect of divided attention. Experiment 2 compared the effects of divided attention at encoding on reporting and monitoring using both non-phased and 2-phase procedures. Divided attention effects were consistent with Experiment 1. However, compared to those in the non-phased condition, participants in the 2-phase condition displayed a more conservative control strategy, and confidence ratings were less diagnostic of accuracy. When memory quality was reduced, although attempts to balance informativeness and accuracy increased the chance of fine-grained response errors, confidence provided an index of the likely accuracy of volunteered fine-grained responses for both condition.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

When individuals remember information about their experiences, they generally also assess the information recalled for likely accuracy. This assessment can be explicit (e.g., an eyewitness might indicate their confidence in the accuracy of a provided detail) or implicit (individuals privately assess the likely accuracy of retrieved information prior to volunteering it; Koriat & Goldsmith, 1996). Explicit expressions of confidence are influential in many real world settings (e.g., police officers, lawyers, and jurors find confident eyewitnesses more compelling than witnesses expressing low confidence). Thus, understanding the confidence–accuracy relation for reported information is a matter of practical importance. However, the extent to which previous investigations of the confidence–accuracy relation speak to memory reporting in applied contexts is limited in important ways (Luna & Martín-Luengo, 2012).

First, detailed examinations of the confidence–accuracy relation for episodic recall are rarely conducted. Research has often focused on

semantic recall, and results from general knowledge domains may not translate to applied episodic recall (e.g., eyewitness recall: Luna & Martín-Luengo, 2012; Perfect, Watson, & Wagstaff, 1993). Although some overlap in the processes supporting episodic and semantic memory is likely, there are dissociable neural correlates for episodic and semantic encoding (Prince, Tsukiura, & Cabeza, 2007), and clinical research demonstrates dissociations in impairment between the two memory systems (Buccione, Fadda, Serra, Caltagirone, & Carlesimo, 2008). Second, previous investigations of recall memory have largely overlooked the potential effects of individuals' implicit assessments of accuracy prior to volunteering responses on the confidence–accuracy relation for those *volunteered* responses.

Individuals reporting information from recall memory face competing demands for informativeness (i.e., provide as much information as possible) and accuracy (i.e., avoid reporting incorrect information). To balance these demands, individuals regulate the *granularity* (i.e., level of detail) of their responses (e.g. Goldsmith, Koriat, & Weinberg-Eliezer, 2002, Weber & Brewer, 2008). Models of strategic memory reporting (Goldsmith et al., 2002; Koriat & Goldsmith, 1996) generally suggest that, when responding to cued recall questions, an individual first attempts to retrieve a highly-detailed, fine-grained

* Corresponding author.

E-mail address: jim.sauer@utas.edu.au (J. Sauer).

response. The probable accuracy of this candidate response is assessed and, if it exceeds a preset criterion, the response is volunteered. If the probable accuracy falls below the criterion, the individual attempts to retrieve a less detailed, coarse-grained response. Although, as noted by Goldsmith et al., the regulation process need not begin with a fine-grained response, the key point is that retrieved information must exceed a criterion level of certainty (and informativeness; Ackerman & Goldsmith, 2008) to be volunteered.

Drawing on this conceptually sophisticated account of memory reporting, we were motivated by a number of outstanding questions. The phased recall paradigm typically used to test this regulation (i.e., Koriat and Goldsmith's (1996) 2-phase approach, described below) forces the generation and evaluation of both fine- and coarse-grained responses. The extent to which this phased format accounts for memory performance in reporting tasks more akin to real-world memory reporting (e.g., where individuals provide unimpeded responses to cued recall questions) merits investigation. Similarly, studies using phased reporting procedures have investigated the confidence–accuracy relationship for *all* items at Phase 1 (i.e., for fine- and coarse-grained candidate answers for each question). In applied settings, the confidence–accuracy relation of interest is for subsets of items that are self-selected for reporting at either fine or coarse levels of granularity. Thus, the extent to which confidence–accuracy relations for fine- and coarse-grained responses reported in previous work extend to non-phased reporting tasks requires scrutiny.

Understanding participants' regulation in more naturalistic reporting tasks is an important (and generally overlooked) aspect of studying the confidence–accuracy relation for episodic recall, particularly as the role of confidence in participants' strategic regulation of memory reporting may have implications for the confidence–accuracy relationship for volunteered responses. Note that the term “volunteered” is typically used in studies of report option, where participants decide whether to offer or withhold an answer. We use the term “volunteered” to distinguish between responses that participants may consider during the regulation process (e.g., at Phase 1 of the 2-phase procedure, or spontaneously during a non-phased procedure) and the responses participants ultimately report. Thus, our interest is not in volunteered versus withheld answers, but in answers volunteered at different levels of granularity.

Using a cued-recall paradigm and manipulating attention at encoding, we investigated how participants' regulation strategy, and the confidence–accuracy relationship for volunteered fine- and coarse-grained episodic memory responses, varied according to memory quality. Our objective was not to test existing accounts of the strategic regulation of grain size, but to extend this literature by (i) determining whether previous examinations of grain size regulation extend to non-phased reporting conditions, particularly when episodic memory quality is reduced through divided attention at encoding (Experiment 1), (ii) exploring the extent to which regulation patterns observed in non-phased reporting conditions map onto previous models (Experiment 2), and (iii) examining how the regulation of grain size affects the confidence–accuracy relation for volunteered information (Experiments 1 and 2).

1.1. Strategic regulation of grain size

Previous research tells us a great deal about participants' ability monitor and control memory outputs, but has focused primarily on the mechanisms underlying, and role of confidence in, the regulation of recall memory output. The 2-phase paradigm was designed for this purpose (see Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008). In the 2-phase method, participants first provide fine- and coarse-grained responses, with associated confidence ratings, for each question asked. Confidence ratings are then removed and participants commit to one response. Intuitively, participants could adopt one of three reporting strategies. First, participants could report only

coarse-grained responses (maximizing accuracy). Second, they could report only fine-grained responses (maximizing informativeness). Finally, they could use a simple “satisficing” strategy: reporting fine-grained responses when confidence exceeds a preset criterion for accuracy and coarse-grained responses when it does not. Research suggests that, although confidence in the initial fine-grained response is the primary determinant of the ultimately volunteered response, the regulation of grain size conforms to a more sophisticated dual-criterion model where participants volunteer fine-grained responses even when confidence is low in order to satisfy a criterion level of informativeness (Ackerman & Goldsmith, 2008; Weber & Brewer, 2008). This strategy holds under conditions of low knowledge, which may offer insight into participants' monitoring and control of weakened episodic memories (Ackerman & Goldsmith, 2008).

Although informative about the processes underlying the strategic regulation of grain size, two methodological elements, common to investigations of grain size regulation, limit the generalizability of previous findings to memory performance in applied settings. First, previous experiments often used methodologies that produce strong memories (e.g., full attention at encoding). Participants may be more likely to prioritize informativeness when memory is strong and fine-grained details are more readily accessible. That is, participants may *be better able* to provide informative (though potentially inaccurate) responses. Under such conditions participants may also infer, from social maxims of communication (Ackerman & Goldsmith, 2008), that they *should* prioritize informative responses. Coarse-grained responses may not seem sufficiently helpful. Second, the 2-phase paradigm itself may affect control processes. In more naturalistic reporting tasks, individuals typically self-regulate the granularity of their responses and report a single response (cf. explicitly outputting fine- and coarse-grained response alternatives before deciding which to volunteer). In the 2-phase paradigm, having generated a fine-grained alternative, participants may be more inclined to volunteer it, or explicit instructions to generate a coarse-grained response may increase the perceived legitimacy of this option. Alternatively, generating both fine- and coarse-grained responses before deciding which to volunteer may simply alter the way participants retrieve and evaluate candidate responses (cf. Perfect & Weber, 2012, on the regulation of report option).

With respect to the strategic regulation of memory, Experiment 1 investigated participants' regulation of response granularity in a non-phased procedure, and under conditions producing sub-optimal memories. Departing from the 2-phased paradigm is necessary to develop our understanding of grain size regulation – and the implications of this regulation for the confidence–accuracy relation – in non-phased reporting tasks. However, although a non-phased approach permits conclusions about the effects of our manipulations on participants' tendency to provide fine- (cf. coarse-) grained responses, the accuracy of these responses, and the diagnostic value of confidence for these responses, it cannot cleanly differentiate the specific monitoring and control processes that lead to these outcomes. To partially attenuate this problem, Experiment 2 includes a direct comparison of the phased and non-phased reporting procedures.

1.2. Memory quality and monitoring

To further explore the effects of memory quality and grain size regulation on episodic memory output we examined participants' *retrospective* metacognitive monitoring as indexed by resolution (i.e., ability to discriminate correct from incorrect responses) and calibration. Importantly, previous research has generally focussed on monitoring *prior* to the volunteering of information (e.g. Ackerman & Goldsmith, 2008, Goldsmith et al., 2002, Weber & Brewer, 2008). In such cases, monitoring is typically discussed as a “competence” (i.e., an individual's ability to discriminate between information that is more or less likely to be correct prior to volunteering it), rather than a source of diagnostic information for someone external to the individual (e.g., a police officer or

juror relying on witness confidence to assess the reliability of a response)¹ and, therefore, a matter of practical significance. Thus, we investigated the effects of memory quality on the diagnostic value of confidence for volunteered responses.

Laboratory experiments rarely replicate the compromised encoding conditions witnesses are likely to experience (Ihlebaek, Love, Eilertsen, & Magnussen, 2003), yet attentional resources at encoding determine the quality of memory for the witnessed event (Lane, 2006). Understanding the confidence–accuracy relationship for volunteered responses is particularly critical when memory quality might have been impaired as a function of divided attention (or related contextual factors) at encoding. Factors that affect episodic memory quality tend to have accompanying effects on metacognitive monitoring for retrieved information (e.g., age-related differences, Perfect & Stollery, 1993; Wong, Cramer, & Gallo, 2012). Variations in memory quality may affect the confidence–accuracy relation for applied recall in two ways. First, although confidence is central to the regulation of memory reporting (Goldsmith et al., 2002; Weber & Brewer, 2008), compared to memory performance itself, confidence is relatively insensitive to changes in memory quality (Gigerenzer, Hoffrage, & Kleinboelting, 1991). Thus, factors that reduce memory quality may impair participants' ability to monitor the accuracy of retrieved responses *prior to* volunteering them (Goldsmith, Koriat, & Pansky, 2005; Perfect & Stollery, 1993), thereby undermining participants' ability to strategically regulate response granularity to satisfy criterion levels of accuracy and informativeness. Second, and the focus of our research, factors that reduce memory quality may impair participants' ability to assess the probable accuracy of *volunteered* responses. This would manifest as impairments in measures of “retrospective monitoring” such as confidence–accuracy calibration and resolution, based on confidence ratings for volunteered responses.

1.2.1. Assessing monitoring pre- and post-volunteering

Using the 2-phase paradigm and testing recall for written testimony, Goldsmith et al. (2005) found that monitoring effectiveness declined with delayed testing. Although their second experiment included a free-grain reporting phase, this free-grain reporting phase was preceded by a forced-fine-grained reporting phase and the confidence–accuracy relationship for responses volunteered in this phase was not reported. Although Koriat and Goldsmith (1996, Experiment 2) concluded that the order of monitoring stages made little difference to the regulation of report option, more recent research (Perfect & Weber, 2012) demonstrates that order can affect monitoring. Thus, we argue that the generalizability of findings obtained using a free-grain reporting *after* forced-fine-grain reporting procedure to conditions where individuals spontaneously regulate grain size should be considered, rather than assumed. Ackerman and Goldsmith (2008, Experiment 1A) included an unimpeded free-grain reporting phase, and demonstrated that participants could balance the demands for informativeness and accuracy for free-grain responses (i.e., that participants could monitor the accuracy of free-grain reporting). They did not, however, assess the confidence–accuracy relation for responses volunteered during this phase, and this study tested semantic, not episodic memory.

Similarly, Kelley and Sahakyan (2003) found that divided attention at encoding impaired monitoring for word pairs. However, they used the 2-phase paradigm and investigated the regulation of report option not grain size. Pansky, Goldsmith, Koriat, and Pearlman-Avni (2009) also found that divided attention impaired monitoring of report option using the 2-phase paradigm for an episodic memory task, but their volunteering phase followed a forced-choice recognition phase which limits the generalizability of their results to recall tasks where

participants must generate and monitor responses. Positive associations have also been found between memory quality and other indices of metacognitive monitoring (e.g., judgments of learning, Koriat, Sheffer, & Ma'yan, 2002; and source monitoring judgments, Lane, 2006), but the relationship between memory and metacognitive performance may vary according to the metacognitive process concerned (Leonesio & Nelson, 1990). Thus, although previous research demonstrates that memory quality affects monitoring prior to volunteering a response, the generalizability of this relationship to applied reporting contexts – when individuals spontaneously regulate grain size for episodic memory reports – requires investigation.

We selected divided attention as our manipulation of memory quality for several reasons. First, the effects of divided attention on memory are well-documented and robust (Iidaka, Anderson, Kapur, Cabeza, & Craik, 2000). Second, the effects of divided attention manipulations at encoding are generalizable across divided attention manipulation types (Fernandes & Moscovitch, 2000). Third, divided attention affects other metacognitive monitoring processes (e.g., regulating report option, source monitoring). Fourth, because divided attention at encoding has larger effects on the recollection of specific details than general familiarity (see Yonelinas, 2002, for a review), it may be particularly pertinent to participants' regulation of response granularity. Finally, the effect of compromised encoding conditions on metacognitive performance is an issue of considerable importance across applied memory domains.

Assessing the reliability of reported information is vital in investigative and, indeed, other applied reporting settings. Thus, our interest is in the within-subjects confidence–accuracy relation: The extent to which participants can adjust their confidence to reflect the likely accuracy of a volunteered response. The limited extant literature for eyewitness recall generally suggests a positive confidence–accuracy relation (e.g. Allwood, Ask, & Granhag, 2005, Robinson & Johnson, 1996). However, the relationship has not received unequivocal support (Perfect, 2002; Roberts & Higham, 2002; Smith, Ellsworth, & Kassir, 1989), the boundary conditions for this relationship remain unclear, and the implications of grain size regulation are unexplored. Although various manipulations weaken the association between (or dissociate) confidence and accuracy for general knowledge questions (e.g., Gigerenzer et al., 1991) and recognition memory (e.g. Busey, Tunnicliff, Loftus, & Loftus, 2000, Weber & Brewer, 2004), research investigating the sensitivity of confidence to changes in memory quality for episodic recall is limited (cf. Odinet and Wolters (2006) who found that increasing retention interval up to five weeks did not significantly impair monitoring for free recall).

Further, witnesses' regulation strategy may have important implications for the confidence–accuracy relationship for responses volunteered at self-selected levels of response granularity. If witnesses continue to favor fine-grained responses when memory quality is reduced (Goldsmith et al., 2005), recall errors are likely to increase (increasing the need for reliable indices of accuracy). If confidence is insensitive to changes in memory quality, monitoring performance may suffer, impairing confidence–accuracy calibration and resolution. Alternatively, increased coarse-grained responding when memory is weak (Goldsmith et al., 2005) may constrain variation in the accuracy of volunteered responses, undermining the confidence–accuracy relationship (cf. Lindsay, Read, & Sharma, 1998). Further, monitoring is poorer for coarse- than fine-grained responses (Goldsmith et al., 2005), thus increased coarse-grained responding may reduce the proportion of responses for which confidence reliably discriminates accuracy. Finally, given that fine-grained responses must exceed some criterion level of confidence in order to be volunteered, the regulation of grain size fundamentally constrains variation in confidence (at least for fine-grained responses), potentially undermining the confidence–accuracy relationship (cf. Lindsay et al., 1998). Such “item selection effects”, considered methodological limitations in basic research, may have serious implications for assessing the reliability of recall responses in applied settings and, as such, are central to our research. In short, the

¹ We thank a previous reviewer for elucidating this distinction. See also Yaniv et al. (1991) for a related consideration of the distinction between discrimination skill and calibration skill.

effects of memory quality and grain size regulation on the confidence–accuracy relationship for volunteered responses demand empirical scrutiny.

Two experiments provide an integrated investigation of the effects of memory quality and grain size regulation on the confidence–accuracy relationship for volunteered responses, investigating three key issues using episodic cued recall paradigms. First, do participants sacrifice accuracy for informativeness in non-phased reporting tasks? Second, if so, to what extent does this compromise the reliability of volunteered responses when memory quality is reduced? Experiment 1 provided an initial answer to these questions by comparing the proportion of cued recall questions eliciting fine- vs. coarse-grained responses, and accuracy for volunteered fine-grained responses in the full vs. divided attention conditions. Including the 2-phase paradigm in Experiment 2 permitted an in-depth analysis of the effect of divided attention on participants' ability to provide correct fine- and coarse-grained answers and, through comparison, facilitated a more informed interpretation of participants' monitoring and control in the non-phased conditions. Finally, we explored the effects of grain size regulation and impaired memory quality on the confidence–accuracy relation for volunteered responses. In both experiments, we addressed this question by comparing confidence–accuracy calibration and resolution in the full vs. divided attention condition.

2. Experiment 1

2.1. Method

2.1.1. Participants

Sixty-four (50 female) undergraduate students and staff, aged 19 to 64 years ($M = 28$, $SD = 10$), participated and received an honorarium for their participation. Participants were allocated randomly to either the divided attention ($n = 30$) or control condition ($n = 34$).

2.1.2. Stimuli

Four color drawings of scenarios varying in detail/scene complexity were taken from the stimuli set used by Gabbert, Memon, Allan, and Wright (2007). These pictures depicted complex scenes, and included both object- and person-related details. For reference, we used the pictures referred to by Gabbert et al. as Kitchen (B), Town Center (A), Living Room (A), and Crossroad (B).

2.1.3. Procedure

Instructions and stimuli were presented by computer. Participants viewed four pictures depicting different scenarios and were asked questions about each picture. Participants were told to respond to questions as if they were eyewitnesses to the scenario presented, and were instructed about the importance of accuracy and informativeness. Instructions² introduced the concept of grain size regulation as a method for balancing the competing goals of accuracy and informativeness, and provided specific examples of fine- and coarse-grained responses (cf. Hope, Gabbert & Brewer, 2010). Participants were told that they should vary the grain size of their responses to “reflect the level of detail they could report accurately”. Participants were informed that they would be asked to provide a retrospective confidence rating for each response. Practice trials ensured that participants understood the task.

Participants completed four blocks of trials, one for each picture. Each block included a 60 sec exposure to the picture, and a 10 sec retention interval (displaying a blank screen) followed by a series of cued-recall questions. During encoding in the divided attention condition, participants listened to randomly ordered numbers spoken at a rate of approx. two numbers per second and were instructed to press the ‘Space’ bar in response to odd numbers.

The number of questions for each picture ranged from 26–29, depending on picture content, with 109 questions in total. Participants answered both number- and color-based questions (Weber & Brewer, 2008). Picture presentation order was randomized and, within blocks, question presentation order was quasi-random. General questions always preceded more targeted questions ensuring participants could not infer any of the answers in each set. Otherwise, question order varied randomly.

Each question reminded participants they could respond with a specific number/color, or a range of values (of a prescribed coarseness)/shade (i.e., light or dark). For example, “What color or shade (e.g., light or dark) was the bus?” and “How many (specific number or range of 3: e.g., 10–12) people were on the bus?”. Participants were not given an explicit *don't know* response option, but could spontaneously respond *don't know* if they wished. Only seven participants (4 in the divided attention condition) provided a *don't know* response, accounting for .42% of all responses. After each response, participants provided retrospective confidence ratings (0–100% scale with decile response options).

2.1.4. Coding responses

Responses were coded for accuracy and grain size. Pre-determined, question-specific parameters defined when responses were considered fine-grained (i.e., indicated the degree of specification required), and the permitted range for coarse-grained responses. Following Weber and Brewer (2008), for color-based questions, responses ‘light’ or ‘dark’ were considered coarse-grain. Specific responses (hot pink, sky blue, etc.) were coded as fine-grain. For number-based questions, the requirements for a fine-grained response varied. For all but two number-based questions, responses were fine-grained if they indicated a single value (e.g., 4) and coarse-grained if indicating a range (e.g., 3 to 8). This approach was appropriate when the correct answer was a quantity that participants could be expected to accurately recall (e.g., three people in a room). However, when relevant quantities were larger (e.g., fifteen rungs on a ladder), the fine- vs. coarse-grained distinction reflected the size of the range indicated. A fine-grained response might include three possible values (e.g., 14–16), whereas a coarse-grained response might include eight possible values (e.g., 12–19). In such cases, permitted ranges for fine- vs. coarse-grained responses were pre-determined and clearly indicated to participants in the recall probe.

2.2. Results

We used two approaches to analyze our data. Where our focus was on trial-by-trial performance (e.g., the effects of our manipulations on response accuracy or the likelihood that a fine- (cf. coarse-) grained response would be volunteered), we used logistic mixed-effects models. Models were created using the `lme4` package (Bates, Maechler, Bolker, & Walker, 2013) in R, an open-source language and environment for statistical computing (R Core Team, 2013). Among other advantages, this approach allowed us to include participant and stimulus as random effects in all models (i.e., allowing random intercepts for these factors), and avoid problems associated with analyzing aggregated means (see Baayen, Davidson, & Bates, 2008; Jaeger, 2008, for a fuller discussion of the advantages of mixed-effects modeling over the standard ANOVA approach). Models analyzing accuracy also included by-participant and by-stimulus random slopes for response granularity. When interpreting these models, as with logistic regression, the outcome variable is the log odds of a particular result. Chi-square tests assess whether the inclusion of a predictor significantly improves the fit of the model, and regression coefficients (b) index the degree of change in the outcome associated with a 1-unit change in the predictor. To aid comprehension, we also present model-estimated probabilities (for fine-grained responding, and response accuracy) according to condition. Although, compared to log odds, probabilities provide a more intuitive visualization of the

² See Supplementary materials.

results, three points must be noted when considering these data. First, these data are provided for descriptive purposes only. We strongly encourage readers to base their interpretations on the coefficients and associated indices of variance provided in the text and relevant tables. Second, one of the advantages of representing outcomes as log odds (cf. probabilities or proportions) is that log odds do not compress outcomes to fit a bounded scale. This compression can distort apparent differences between conditions, and becomes greater as the observed value departs from log odds of 0 (i.e., proportion of .50). Thus, differences expressed in log odds and probabilistic terms will not necessarily appear equivalent. Finally, the values provided are based on model estimated probabilities and exclude effects associated with random factors: They are not mean proportions. For the purpose of comparison with previous research, we include traditional ANOVA-based analyses of raw data in the supplementary materials. When measures were based on samples of responses (e.g., the adjusted normalized discrimination index, *ANDI*, see Yaniv, Yates, & Smith, 1991), multi-level modeling was not appropriate and we used standard ANOVA analyses. We used Cohen's *f* to index effect size. Cut-off values for small, moderate, and large effects are 0.10, 0.25, and 0.40, respectively.

2.2.1. Memory quality, response granularity and accuracy

We sought to determine if participants sacrificed accuracy for informativeness when memory quality was reduced. Evidence of participants' willingness to sacrifice accuracy for informativeness would be provided if (a) the proportion of items receiving fine-grained responses was unaffected by the attention manipulation and/or fine-grained response accuracy was lower in the divided- than full-attention condition, and (b) this difference was not attributable to a reduction in participants' ability to discriminate correct from incorrect responses (indexed by *ANDI*). Lower mean confidence for fine-grained responses in the divided- (cf. full attention conditions) would be consistent with a more lenient criterion for providing fine-grained responses, and also suggest a willingness to prioritize informativeness over accuracy.

To test if the divided attention manipulation affected the proportion of items receiving fine-grained responses we constructed a logistic mixed-effects model with response granularity (fine-grained or coarse-grained) as the outcome variable and participant (*SD* = 0.90) and stimulus (*SD* = 0.33) as random effects. Adding the attention component significantly improved the fit of the model, $\chi^2(1) = 10.67, p = .001$, and indicated that participants were less likely to provide fine-grained responses when attention had been divided (cf. full) at encoding, $b = -.83, SE_b = 0.24, p < .001$. This equates to the probability of a volunteering a fine-grained response being .74 and .56 for the full and divided-attention conditions, respectively. To rule out the possibility that this reduction in fine-grained responding indicated an effort to prioritize accuracy, we tested if the reduction in fine-grained responses attenuated any deleterious effects of divided attention at encoding on the accuracy of volunteered fine-grained responses. We constructed another logistic mixed-effects model with accuracy as the outcome variable, and participant (*SD* = 0.40) and stimulus (*SD* = 0.65) as random effects. Adding attention significantly improved the fit of the model, $\chi^2(1) = 4.85, p = .028$, as did the attention \times response granularity interaction, $\chi^2(1) = 9.62, p = .001$ (see Table 1 for model coefficients). As can be seen in Fig. 1 (which presents accuracy in terms of both predicted log odds and model-estimated probabilities),

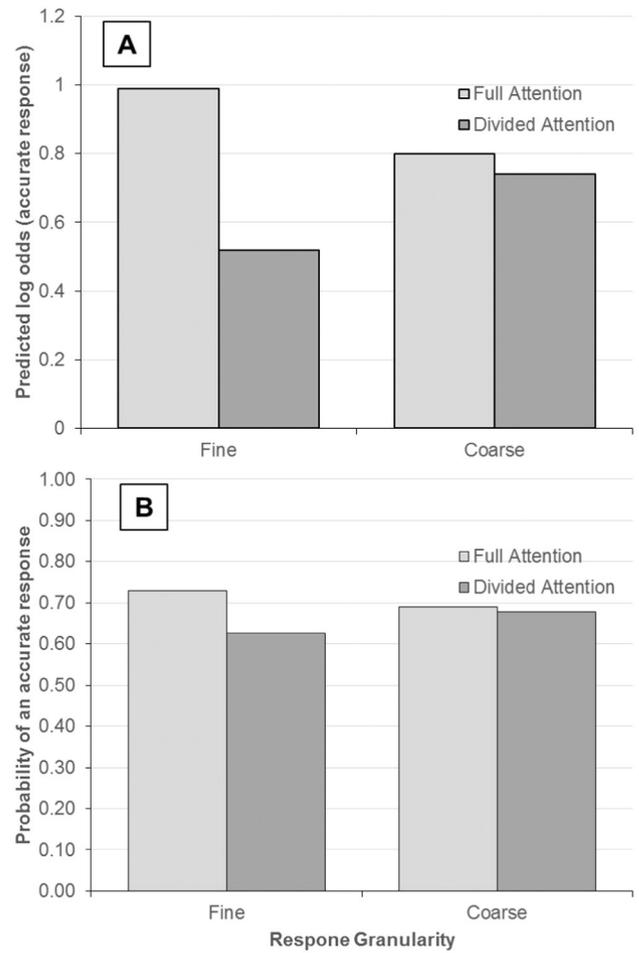


Fig. 1. The predicted log odds (Panel A) and model-estimated probability (Panel B) of an accurate response according to attention and response granularity (Experiment 1). Coefficient standard errors, indicating estimate variability for the log odds, are provided in Table 1.

participants whose attention was divided (cf. full) at encoding had lower fine-grained response accuracy, despite providing fewer fine-grained responses. An inspection of differences in fine-grained response confidence between the divided and full attention conditions can also provide some insight into participants' control processes. Fine-grained response confidence has been identified as the primary determinant in the decision to volunteer a fine- vs. coarse-grained responses, such that a fine-grained candidate response must exceed a criterion level of subjective accuracy (i.e., confidence) in order to be volunteered (Goldsmith et al., 2002; Weber & Brewer, 2008). If participants relaxed their accuracy criterion to sacrifice accuracy for informativeness (when memory quality was relatively weak), we would expect to see lower mean confidence for volunteered fine-grained responses in the divided (cf. full) attention condition.³ As expected, confidence for volunteered fine-grained responses was significantly lower in the divided ($M = .70, SD = .15, [.65, .76]$) than the full attention condition ($M = .85, SD = .07, [.83, .87]$), $t(62) = 5.32, p = .001, f = 0.61$. To maintain levels of informativeness and compensate for their relatively poor memory quality, participants in the divided attention condition provided fine-grained responses with lower subjective accuracy.

2.2.2. Memory quality and the confidence–accuracy relation

First, we analyzed participants' ability to discriminate between retrieved information that is more or less likely to be reliable (using the

Table 1
Fixed effect coefficients for logistic mixed-effects model predicting accuracy (Experiment 1).

Fixed effect	<i>b</i>	<i>SE_b</i>	<i>Z</i>	<i>p</i>
Intercept	0.52	0.34	1.52	.128
Attention (A)	0.47	0.12	3.80	<.001
Response Granularity (RG)	0.22	0.34	0.65	.519
A \times RG	−0.41	0.12	3.30	<.001

³ We thank a reviewer for suggesting this analysis.

ANDI statistic). These analyses specifically investigate whether the lower fine-grained response accuracy in the divided (cf. full) attention condition can be attributed to participants' inability to discriminate correct from incorrect responses (as opposed to strategic regulation), and speak to participants' regulation of response granularity in the non-phased task. Second, we present calibration curves which compare the objective and subjective likelihood of response accuracy, and speak the utility of confidence as an index of accuracy for volunteered responses. Third, we investigated the effects of our attention manipulation and response granularity on participants over/underconfidence (OU; sometimes referred to as *calibration bias*).

As suggested above, the observed reduction in fine-grained accuracy may indicate an effect of our manipulation on metacognitive monitoring, rather than memory quality or participants' willingness to sacrifice accuracy for informativeness. That is, participants may have attempted to prioritize accuracy but been unsuccessful due to impaired monitoring ability in the divided attention condition. The difference between conditions in the mean confidence for volunteered fine-grained responses suggests this is not the case but, to test this idea, the effects of divided attention on participants' metacognitive monitoring (ANDI) were assessed⁴. Differences between conditions in metacognitive monitoring for fine-grained responses might implicate impaired monitoring (rather than strategic regulation) in the observed reduction in fine-grained response accuracy.

ANDI ranges from 0 (*no discrimination*) to 1 (*perfect discrimination*), and measures retrospective monitoring ability (i.e., the extent to which confidence ratings discriminated correct from incorrect responses; see Table 2). Specifically, ANDI indexes the amount of variance in accuracy explained by participants' confidence ratings, and can be interpreted as eta-squared from a one-way ANOVA. ANDI has two primary advantages over other measures of discrimination (e.g., Goodman and Kruskal's (1954) Gamma coefficient). First, ANDI normalizes variance explained according to baseline levels of variance in outcome measure, thereby removing a statistical ceiling from achievable levels of discrimination. Second, ANDI is unaffected by the number of response categories or the number of judgments made in each response category, making it a bias-free measure of discrimination (Yaniv et al., 1991). To counter skewness, a square root transformation was applied to the mean ANDI data ($ANDI_{sqr}$). In all conditions, mean $ANDI_{sqr}$ statistics were significantly greater than zero indicating that confidence discriminated correct from incorrect responses. Moreover, a 2 (attention: divided vs. full) \times 2 (response granularity: fine vs. coarse) mixed ANOVA on mean $ANDI_{sqr}$ (see Table 3) returned a non-significant main effect of attention, and a non-significant Attention \times Response Granularity interaction. Both these effects failed to reach the cut-off for a small effect and, thus, there is no evidence that divided attention at encoding affected participants' retrospective monitoring, or that the attention effect on fine-grained response accuracy was attributable to impaired monitoring. Of course, a non-significant difference is not the same as equivalence. Therefore, we ran a Bayes Factor analysis comparing $ANDI_{sqr}$ for fine-grained responses in the divided- and full-attention conditions (Morey & Rouder, 2014; Rouder, Speckman, Sun, Morey, & Iverson, 2009). Bayes Factors index the relative evidence, given the obtained data, supporting the null and alternative hypotheses. The $JZS BF_{01} = 3.30$, indicating that the data were approximately 3.30 times more likely under the null than the alternative hypothesis. According to Jeffreys (1961), this represents substantial evidence in favor of the null hypothesis. Finally, there was a significant main effect of grain size. Consistent with the calibration curves, discrimination was better for fine- than coarse-grained responses.

Second, calibration curves compared the objective and subjective probabilities of accuracy for fine- and coarse-grained responses in the full and divided attention conditions (see Fig. 2 for the curves and

Table 4 for response frequencies at each confidence level). The curves for fine-grained responses show linear, positive relationships between confidence and accuracy for both conditions. Thus, in both conditions, participants' confidence is diagnostic of fine-grained response accuracy. The coarse-grained curves show no evidence of a meaningful relationship in either condition. The OU statistic assesses the extent to which, on average, participants' confidence ratings were greater or less than their accuracy. OU ranges from -1 to $+1$, with positive and negative scores indicating over- and underconfidence, respectively. Data screening identified one outlier (greater than 3.29 SDs from the mean: Tabachnick & Fidell, 2007) for each of the fine- and coarse-grained mean OU data distributions. These outliers were removed prior to analysis. A 2 (attention: divided vs. full) \times 2 (response granularity: fine vs. coarse) mixed ANOVA on mean OU returned a significant main effect of response granularity, $F(1, 60) = 184.11, p < .001, f = 0.86$: Participants were generally overconfident for fine-grained responses ($M = .09, SD = .10, 95\% CI [.07, .12]$) but underconfident for coarse-grained responses ($M = -.14, SD = .17, [-.18, -.10]$). The main effect of attention was non-significant though non-trivial, $F(1, 60) = 2.28, p = .137, f = 0.18$, and suggested that divided (cf. full) attention may have been associated with slightly increased underconfidence ($M = .00, SD = .19, [-.05, .05]$, and $M = -.05, SD = .18, [-.10, .00]$, for the full and divided attention conditions, respectively). However, looking at the obtained calibration curves, there appears to be little systematic variation in the realism of confidence ratings between conditions for either fine- or coarse-grained responses. The interaction was trivial, and non-significant, $F(1, 60) = 0.09, p = .768, f = 0.02$.

2.3. Discussion

Experiment 1 investigated (i) participants' regulation of response granularity under non-phased reporting conditions; (ii) how regulation was affected by memory quality, and (iii) how regulation affected the confidence-accuracy relation for volunteered responses. Participants appear to have sacrificed accuracy for informativeness, even when memory was relatively weak. Participants in the divided (cf. full) attention condition provided fewer fine-grained responses (consistent with Goldsmith et al., 2005), but this did not prevent a reduction in the accuracy of fine-grained responses. Although the non-phased paradigm cannot cleanly differentiate monitoring from control effects on outputs, the lower mean confidence for fine-grained responses in the divided (cf. full) attention condition together with the absence of any effect of memory quality on participants' ability to discriminate correct from incorrect fine-grained responses suggests that this pattern indicates a willingness to sacrifice accuracy in favor of informativeness (i.e., a control effect), rather than an inability to monitor the accuracy of fine-grained information. This finding appears consistent with research using the 2-phase paradigm (e.g. Ackerman & Goldsmith, 2008, Weber & Brewer, 2008). However, using the non-phased paradigm, we cannot rule out the possibility that the lower mean confidence for fine-grained responses reflects an inability to answer questions with higher confidence, even at the coarse-grained level. Experiment 2 addresses this issue with a direct comparison of the non-phased and 2-phase procedures.

We found no effect of divided attention on resolution and no visible difference in the linearity of the calibration curves for reported information. Thus, we found no evidence that reduced memory quality adversely affected of the confidence-accuracy relationship for volunteered responses. Given the applied value of discriminating between memory reports that are more or less likely to be reliable, this finding is encouraging.

3. Experiment 2

Experiment 1 provided an initial investigation of the effect of divided attention at encoding on participants' regulation of response granularity, and the confidence-accuracy relationship for volunteered responses, using a non-phased paradigm. However, this investigation did not

⁴ For the purposes of comparison with previous work in the literature, analyses based on gamma coefficients can be found in the Supplementary materials.

Table 2

Mean, SD, and 95% CI square root transformed ANDI statistics for fine- and coarse-grained responses for participants in the full and divided attention conditions (Experiment 1).

Attention & response granularity	ANDI _{sqrt}		
	M	SD	95% CI
Fine-grained			
Full attention	.46	.17	.40–.51
Divided attention	.49	.14	.44–.54
Overall	.47	.16	.43–.51
Coarse-grained			
Full attention	.21	.23	.12–.29
Divided attention	.20	.20	.13–.28
Overall	.20	.22	.15–.26

permit an in-depth examination of participants' *regulation strategy*. Specifically, without being able to examine the accuracy (and confidence) of both fine- and coarse-grained response alternatives, we could not directly investigate if the observed effects were attributable to differences in monitoring (i.e., an ability deficit), control (i.e., a strategic choice), or actual knowledge. Thus, our primary aim for Experiment 2 was to establish if the findings from Experiment 1 could be replicated and, assuming they would be, use a comparison with a 2-phase reporting procedure to determine if the effect of divided attention on fine-grained accuracy in the non-phased condition indicated the absence of knowledge or a willingness to sacrifice accuracy for informativeness.

Including a direct comparison of the non-phased and 2-phase procedures allowed us to answer another important question: Does the 2-phase paradigm itself affect participants' monitoring and control processes? This question is of theoretical and applied importance. From a theoretical perspective, there are reasons to expect that reporting procedure may influence participants' regulation strategy and responding. If a direct comparison of the two procedures reveals systematic differences, it may also offer some insight into the effects of reporting procedure on underlying metacognitive processes or, at least, highlight an important avenue for future investigation. From an applied perspective, it is important to determine if the results obtained by grain size regulation research using the 2-phase paradigm can be generalized to non-phased reporting conditions. It is worth noting that the 2-phase paradigm was not intended to approximate reporting conditions that typically interest applied memory researchers. Instead, it was intended to elucidate the metacognitive processes underlying the strategic regulation of memory outputs. However, given the obvious relevance of strategic regulation in applied memory settings and in the absence of data speaking to strategic regulation in non-phased reporting tasks, researchers may understandably use data obtained using the 2-phase paradigm to draw inferences about the regulation of response granularity in more naturalistic settings. Thus, an important aim of Experiment 2 was to establish how well these results generalize, and how appropriate these inferences are likely to be.

Table 3

Mixed ANOVA on square root transformed ANDI statistics (Experiment 1).

Effect	df	F	f	p
Between-subjects				
Attention (A)	1	0.20	0.05	.653
A error	60	(0.01)		
Within-subjects				
Response granularity (RG)	1	62.11	0.71	.000
RG × A	1	0.31	0.06	.580
RG error	60	(0.04)		

Note: Values in parentheses represent mean-square errors.

3.1. Method

3.1.1. Participants and design

We used a 2 (Attention: full vs. divided) × 2 (Report procedure: non-phased vs. 2-phase), between-participants design. Eighty (63 female) undergraduate students and community members, aged 16 to 55 years ($M = 23, SD = 5$), participated and received course credit (students) or chocolate (community members) for their participation. Participants were randomly allocated across cells ($n = 20$ in each cell).

3.1.2. Stimuli

The stimuli were identical to those used in Experiment 1.

3.1.3. Procedure

The procedure for participants in the non-phased condition was identical to that for Experiment 1. Following Weber and Brewer (2008), participants in the 2-phase condition, at Phase 1, provided a fine- and coarse-grained response for each question, and a retrospective confidence rating for each response. At Phase 2, these confidence ratings were removed and, for each question, participants selected one of the available responses to volunteer, and provided a retrospective confidence rating for the volunteered response. As per Experiment 1, participants were not given an explicit *don't know* response option, but could spontaneously respond *don't know*. Again, very few participants provided *don't know* responses. Three participants in the non-phased conditions (one in the divided attention condition) provided at least one *don't know* response, accounting for .3% of all responses in these conditions. No participants in the 2-phase conditions provided a *don't know* response.

3.1.4. Coding responses

Responses were coded for accuracy and grain size, as per Experiment 1. For the 2-phased procedure, our analyses of confidence were based on the confidence rating provided at Phase 2 (i.e., for the volunteered response).

3.2. Results

As per Experiment 1, we used a combination of logistic mixed-effects models, including participant and stimulus as random effects and by-participant and by-stimulus random intercepts for response granularity, and standard ANOVAs to analyze results. Mixed-effect model outputs are presented in log odds and, to aid comprehension, probabilities.

3.2.1. Memory quality, response granularity and accuracy

For each analysis, we first determined if the findings from Experiment 1 were replicated. None of the observed effects of reporting procedure qualify our conclusions relating to this primary aim. Thus, in the interest of conceptual clarity, interpretation of effects relating to reporting procedure is generally provided in a separate section (see *The effects of reporting procedure on monitoring and accuracy*).

As per Experiment 1, to determine if participants sacrificed accuracy for informativeness when memory quality was reduced we first constructed a logistic mixed-effects model with response granularity (fine-grained or coarse-grained) as the outcome variable and participant ($SD = 0.52$) and stimulus ($SD = 0.40$) as random effects (see Table 5 for model coefficients). The addition of attention, $\chi^2(1) = 17.33, p < .001$, and reporting procedure, $\chi^2(1) = 44.72, p < .001$, significantly improved the fit of the model. As can be seen in Fig. 3, and consistent with Experiment 1, participants were more likely to provide fine-grained in the full (cf. divided) attention conditions. The non-significant Attention × Reporting Procedure interaction indicated no evidence that the effect of divided attention on the proportion of questions receiving fine-grained responses was driven by participants in the 2-phase procedure.

Again, we tested whether this reduction in fine-grained reporting in the divided attention condition ameliorated any effect of divided

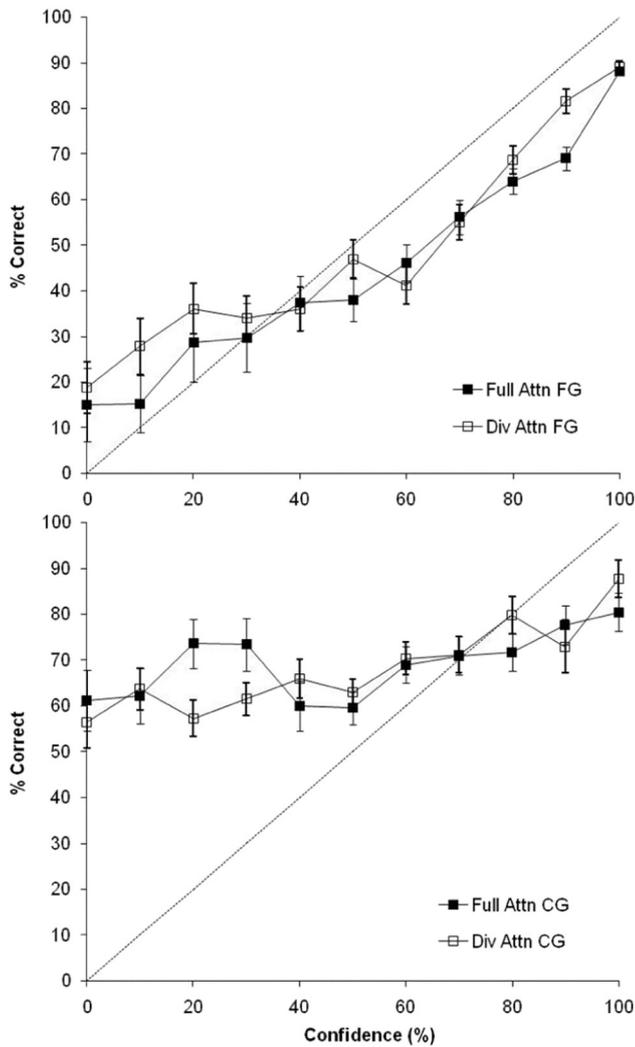


Fig. 2. Calibration curves, based on data collapsed across participants, for volunteered fine-grained (upper panel) and coarse-grained (lower panel) responses in the divided and full attention conditions (Experiment 1). Error bars represent standard errors.

attention on the accuracy of volunteered fine-grained responses. We constructed another logistic mixed-effects model with accuracy as the outcome variable, and participant ($SD = 0.13$) and stimulus ($SD = 0.22$) as random effects (see Table 6 for model coefficients). Although adding attention did not significantly improve the fit of the model, $\chi^2(1) = 0.57, p = .450$, adding the Attention \times Response Granularity interaction did, $\chi^2(1) = 26.00, p < .001$. Consistent with Experiment 1, divided attention at encoding reduced accuracy for fine-grained responses (see Fig. 4). Again, despite volunteering fewer fine-grained responses, participants in the divided (cf. full) attention condition still showed lower fine-grained response accuracy. Adding reporting procedure improved the fit of the model, $\chi^2(1) = 16.90, p < .001$, and indicated higher accuracy in the 2-phase (cf. non-phased) condition. This demonstrates that participants in the non-phased condition had

sufficient knowledge to permit greater accuracy, and that the effects reflect the regulation of response granularity. The Attention \times Response Granularity \times Reporting Procedure interaction did not significantly improve the fit of the model, $\chi^2(1) = 1.82, p = .177$, again indicating no evidence that the effect of attention on fine-grained accuracy varied according to reporting procedure. Thus, in both conditions, results are consistent with participants' sacrificing accuracy for informativeness, even when memory was relatively weak.

Further evidence for this conclusion was provided by reduced mean confidence for volunteered fine-grained responses in the divided (cf. full) attention condition. A 2 (attention: divided vs. full) \times 2 (reporting procedure: non-phased vs. 2-phase) between-subjects ANOVA on confidence for volunteered fine-grained responses found a large main effect of attention, $F(1, 76) = 12.23, p = .001, f = 0.40$. Consistent with a willingness to sacrifice accuracy for informativeness, confidence for volunteered fine-grained responses was significantly lower in the divided ($M = .73, SD = .16, [.68, .78]$) than the full attention condition ($M = .82, SD = .13, [.78, .86]$). Importantly for the replication of findings from Experiment 1, the small but non-significant Attention \times Reporting Procedure interaction, $F(1, 76) = 0.94, p = .335, f = 0.11$, indicated that the main effect of divided attention on fine-grained response confidence was not driven by effects in the 2-phase condition. In fact, if anything the effect was stronger in the non-phased (cf. 2-phase) procedure. The large main effect of reporting procedure, $F(1, 76) = 39.89, p < .001, f = 0.72$, demonstrated that, compared to participants in the non-phased procedure ($M = .70, SD = .14, [.65, .74]$), participants the 2-phase conditions were generally more conservative in volunteering fine-grained responses ($M = .86, SD = .10, [.83, .89]$).

3.2.1.1. The effects of reporting procedure on response granularity and accuracy. The pattern of results for the non-phased and 2-phase conditions suggested that the reporting procedure used affected participants' responding. Specifically, these results are consistent with a relatively conservative regulation strategy in the 2-phase condition. First, participants volunteered fewer fine-grained responses in the 2-phase (cf. non-phased) condition (see Fig. 3). Second, mean confidence for volunteered fine-grained responses was higher in the 2-phase (cf. non-phased) condition. Third, the accuracy of volunteered fine-grained responses was higher in the 2-phase (cf. non-phased) condition (see Fig. 4).

3.2.2. Memory quality and the confidence–accuracy relation

As per Experiment 1, to conclude that the observed effects in the non-phased condition indicated a willingness to sacrifice accuracy, we needed to rule out the possibility that the lower accuracy for non-phased fine-grained responses in the divided (cf. full) attention condition was due to impaired discrimination. Three aspects of the results speak against an impaired discrimination account. First, in all conditions, mean AND_{sqrt} (transformed to counter skewness) statistics were significantly greater than zero, indicating that confidence discriminated correct from incorrect responses (see Tables 7 and 8 for descriptive and inferential statistics, respectively). Second, and more importantly, an inspection of the mean AND_{sqrt} statistics for the non-phased condition suggests that the divided attention manipulation did not impair discrimination (DA $M = .50$; FA $M = .41$). Thus, consistent with Experiment 1, there is no evidence that the lower fine-grained response accuracy in the divided (cf. full) attention non-phased condition

Table 4
Response frequencies for each confidence level according to response granularity and attention (Experiment 1).

Grain size & attention		Confidence level (%)										
		0	10	20	30	40	50	60	70	80	90	100
Fine-grained	Full	20	33	28	37	67	108	150	180	297	316	1350
	Divided	48	54	75	97	100	145	141	171	220	216	541
Coarse-grained	Full	54	66	68	60	80	195	135	117	127	98	92
	Divided	78	110	159	187	126	280	165	128	99	59	65

Table 5

Fixed effect coefficients for logistic mixed-effects model predicting response granularity (Experiment 2).

Fixed effect	<i>b</i>	SE _{<i>b</i>}	<i>Z</i>	<i>p</i>
Intercept	0.56	0.35	1.61	.108
Attention (A)	0.99	0.17	5.85	<.001
Reporting Procedure (RP)	−1.30	0.17	7.69	<.001

resulted from impaired discrimination. Finally, again consistent with Experiment 1, the calibration curves for fine-grained responses in the non-phased condition (see Fig. 5 for the curves and Table 9 for response frequencies at each confidence level) show linear, positive relationships between confidence and accuracy for both the full and divided attention conditions (indicating a robust confidence–accuracy relation for volunteered responses)⁵.

Again, when looking at over/underconfidence, data screening identified one outlier – in the fine-grained mean OU data distribution – and this outlier was removed prior to analysis. A 2 (attention: divided vs. full) × 2 (response granularity: fine vs. coarse) × 2 (reporting procedure: non-phased vs. 2-phase) mixed ANOVA on mean OU returned a significant main effect of response granularity, $F(1, 75) = 22.78, p < .001, f = 0.30$, with fine-grained ($M = .05, SD = .11, [.03, .07]$) and coarse-grained responses ($M = -.04, SD = .21, [−.09, .01]$) showing over- and underconfidence, respectively. The main effect of attention was significant, $F(1, 75) = 5.11, p = .027, f = 0.23$, however this should be interpreted in the context of the significant Attention × Response Granularity interaction, $F(1, 75) = 16.09, p < .001, f = 0.25$. Follow-up Bonferroni-corrected simple effects analyses revealed that divided attention increased underconfidence for coarse-grained responses ($M = .03, SD = .20, [−.03, .09]$ and $M = -.12, SD = .19, [−.18, −.06]$ for the full and divided attention conditions, respectively), $t(78) = 3.47, p = .001, f = 0.39$, but had no significant effect on over/underconfidence for fine-grained responses ($M = .05, SD = .11, [.02, .08]$ and $M = .05, SD = .11, [.02, .08]$ for the full and divided attention conditions, respectively), $t(77) = 0.28, p = .781, f = 0.03$.

Given the effects of the divided attention manipulation on accuracy, and the well-established hard/easy effect (i.e., showing that overconfidence increases with item difficulty), we might have predicted that divided (cf. full) attention would increase overconfidence, rather than underconfidence. Two explanations for this unanticipated pattern of results present themselves. First, participants might have some insight into the effects divided attention on accuracy, and may have overcompensated for this when adjusting their confidence ratings. Second, and perhaps more likely, this pattern might represent an item selection artefact. Participants' might choose to volunteer coarse-grained responses only when the available evidence will not support a fine-grained response. The general underconfidence observed for coarse-grained responses might reflect a paucity of the sort of evidence that drives confidence, and this may be exaggerated when attention was divided at encoding (leading to increased underconfidence). We return to this issue in the general discussion. All other effects were non-significant, $F < 0.69, p$'s $> .408, f$'s < 0.09 .

3.2.2.1. The effects of reporting procedure on the confidence–accuracy relation. There is evidence that reporting procedure affected discrimination and the diagnostic value of participants' confidence ratings. First, although the 2-phase calibration curves for fine-grained responses suggest a monotonic, positive confidence–accuracy relation in the upper half of the confidence scale, a 2 (attention: divided vs. full) × 2 (response granularity: fine vs. coarse) × 2 (reporting procedure) mixed ANOVA on mean within-subjects $ANDI_{sqrt}$ (Table 8) revealed a

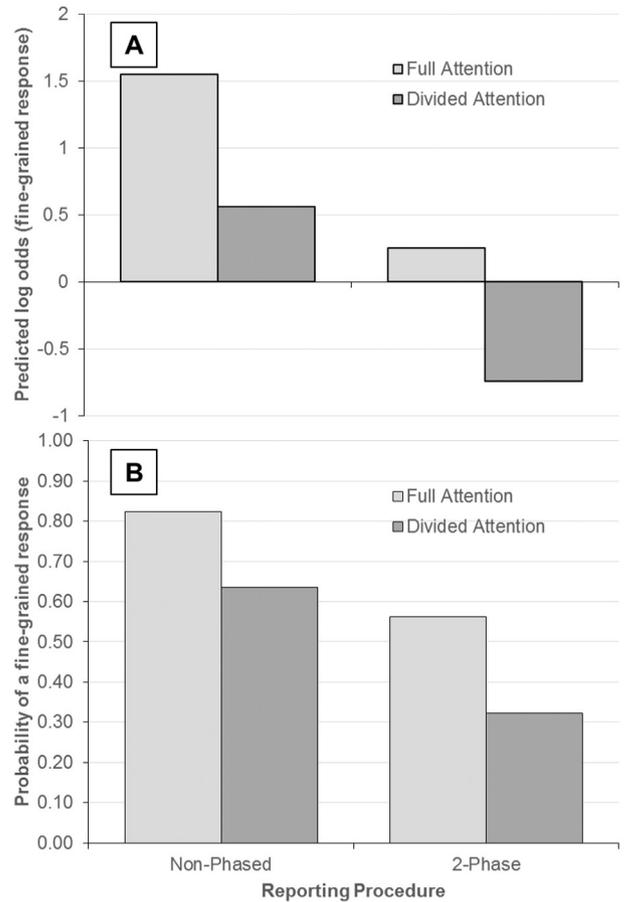


Fig. 3. The predicted log odds (Panel A) and model-estimated probability (Panel B) of a fine-grained response according to attention and reporting procedure (Experiment 2). Coefficient standard errors, indicating estimate variability, are provided in Table 4.

significant main effect of reporting procedure, indicating better discrimination for the non-phased than 2-phase procedure (Table 7). Further, a significant Response Granularity × Attention × Reporting Procedure interaction, together with Bonferroni-corrected simple effects analyses, revealed that divided (cf. full) attention at encoding significantly reduced mean $ANDI_{sqrt}$ for fine-grained responses in the 2-phase condition, $t(38) = 2.77, p = .009, f = 0.45$, but not the non-phased condition, $t(38) = 1.82, p = .077, f = 0.29$. Note that, as indicated above, although the effect for the non-phased condition approaches significance, this actually suggests improved discrimination in the divided (cf. full) attention condition⁶.

These results appear to reflect a relatively strict criterion for volunteering fine-grained responses (consistent with analyses presented above) and, consequently, constrained variation in confidence and accuracy, for fine-grained responses in the 2-phase conditions. Together with the observed effects on grain size regulation and response accuracy, these effects represent a pattern of findings suggesting difficult conditions exaggerate the tendency of participants using the 2-phase (cf. non-phased) procedure to be more conservative in volunteering fine-grained responses, constraining variation in confidence and accuracy for fine-grained responses, and undermining discrimination. Note, however, that this did not undermine confidence–accuracy calibration for volunteered fine-grained responses (Fig. 5).

⁵ Fig. 5 also displays calibration curves for coarse-grained responses. As in Experiment 1, coarse-grained response curves show no evidence of a meaningful relationship in any condition.

⁶ For analyses of gamma coefficients, see Supplementary materials.

Table 6
Fixed effect coefficients for logistic mixed-effects model predicting accuracy (Experiment 2).

Fixed effect	<i>b</i>	SE _{<i>b</i>}	<i>Z</i>	<i>p</i>
Intercept	0.42	0.31	1.36	.174
Attention (A)	0.53	0.12	4.35	<.001
Reporting Procedure (RP)	0.58	0.12	4.71	<.001
Response granularity (RG)	−0.12	0.26	0.46	.648
A × RG	−0.74	0.13	5.65	<.001
RP × RG	−0.31	0.13	2.37	.018

3.3. Discussion

Our primary aims in Experiment 2 were to (i) replicate Experiment 1, and (ii) examine whether the observed effects indicated the absence of knowledge or a willingness to sacrifice accuracy in favor of informativeness. The results for the non-phased condition in Experiment 1 were replicated. Moreover, a comparison with the accuracy rates observed for fine-grained responses provided by participants in the 2-phase procedure indicated that participants had the knowledge to permit greater accuracy, but sacrificed accuracy for informativeness. Our secondary aim was to provide a direct comparison of the non-phased and 2-phase procedures, and test the extent to which results obtained using the 2-phase procedure can be generalized to non-phased reporting conditions. In addition to elucidating the mechanisms underlying the results observed for the non-phased condition, the differences observed between the non-phased and 2-phase reporting conditions supported our hypothesis that the regulation procedure used may affect participants' memory reports. Specifically, in this experiment, the 2-phase regulation condition led to more conservative responding.

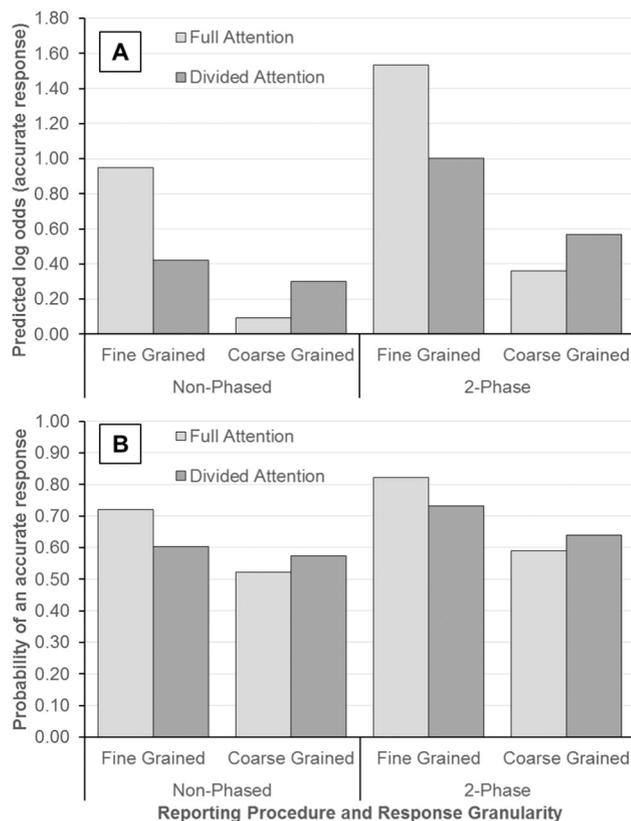


Fig. 4. The predicted log odds (Panel A) and model-estimated probability (Panel B) of an accurate response according to attention, reporting procedure and response granularity (Experiment 2). Coefficient standard errors, indicating estimate variability, are provided in Table 5.

Participants in the 2-phase (cf. non-phased) conditions volunteered fewer fine-grained responses, leading to higher accuracy for volunteered fine-grained responses. This conservative responding strategy impaired participants' discrimination for volunteered fine-grained responses when memory quality was weak, but not their calibration.

4. General discussion

This research provides an integrated investigation of the effect of reduced memory quality on the regulation of grain size and the confidence–accuracy relationship for recall responses, and the first direct comparison of the more typically-used 2-phase paradigm with a non-phased reporting and regulation procedure. Consistent with Ackerman and Goldsmith's (2008) findings for a semantic memory task, participants in our non-phased episodic memory experiment sacrificed accuracy for informativeness, even when memory quality was weak, leading to reduced accuracy for volunteered fine-grained responses in the divided (cf. full) attention conditions. However, participants maintained their ability to monitor the accuracy of volunteered fine-grained responses even when memory quality was reduced, and confidence for these responses remained diagnostic of accuracy. This finding runs counter to Pansky et al.'s (2009) finding that divided attention at encoding impaired monitoring for episodic memories at test. This apparent discrepancy may be attributable to a variety of differences between the two pieces of research. First, the first phase of Pansky et al.'s procedure involved a 5-alternative, forced choice recognition task that differs from our cued-recall task. Second, Pansky et al. included an accuracy incentive, previously shown to influence the regulation of report option (Koriat & Goldsmith, 1996) and grain-size (Goldsmith et al., 2002). Third, Pansky et al. tested the regulation of report option (the choice to volunteer vs. withhold information) rather than the regulation of response granularity. Finally, the confidence ratings used to assess monitoring were provided for the Phase 1 forced-choice recognition responses, rather than Phase 2 (volunteered) responses. Regardless, from an applied perspective, the absence of a memory quality effect on monitoring for volunteered responses in our work is a positive finding.

Our secondary goal, in Experiment 2, was to compare grain size regulation, response accuracy and the confidence–accuracy relationship for volunteered responses for participants using the non-phased and 2-phase reporting paradigms. Differences between the non-phased and 2-phase conditions highlight a number of issues for consideration. Most obviously, while the 2-phase procedure allows in-depth analysis of participants' monitoring and control processes, the generalizability of results obtained using this phased responding paradigm to non-phased reporting conditions is not guaranteed. Our findings suggest that the regulation procedure used exerted systematic effects on participants' responding, accuracy, and discrimination for volunteered responses. Specifically, participants in the 2-phase condition were more conservative in their regulation of grain size than those in the non-phased condition, and this relative difference was exaggerated when memory quality was reduced. Of course, the 2-phase paradigm was designed to investigate the metacognitive processes underlying the strategic regulation of memory outputs, rather than estimate memory reporting in an absolute sense. However, because strategic regulation is clearly relevant in applied memory settings, researchers may understandably generalize from data obtained using the 2-phase paradigm to more naturalistic reporting settings. Investigating the generalizability of these data is therefore a matter of considerable importance and the current research provides an initial account of the differences emerging as a function of regulation procedure.

Confidence reliably discriminated the accuracy of non-phased fine-grained responses but, in contrast to research using the 2-phase paradigm and testing monitoring at Phase 1 of the process (e.g. Goldsmith et al., 2005; Weber & Brewer, 2008), discrimination was poor for volunteered coarse-grained responses. This finding underscores the difference between coarse examining the confidence–accuracy relation for fine-

Table 7

Mean, SD, and 95% CI square root transformed ANDI for fine- and coarse-grained responses for participants in the full and divided attention conditions, using the non-phased and 2-phase reporting procedures (Experiment 2).

Attention & response granularity	ANDI _{sqrt}		
	M	SD	95% CI
Non-phased			
Fine-grained			
Full attention	.41	.16	.33–.48
Divided attention	.50	.11	.44–.55
Overall	.45	.14	.41–.50
Coarse-grained			
Full attention	.23	.27	.10–.36
Divided attention	.13	.16	.06–.21
Overall	.18	.23	.11–.25
2-phase			
Fine-grained			
Full attention	.33	.24	.22–.45
Divided attention	.18	.18	.10–.26
Overall	.26	.22	.19–.33
Coarse-grained			
Full attention	.08	.13	.01–.14
Divided attention	.13	.13	.07–.19
Overall	.10	.13	.06–.15

and coarse-grained responses for all items, and the confidence–accuracy relation for subsets of items self-selected for reporting at different levels of response granularity. This observation has theoretical, applied, and methodological implications. The regulation of grain size is an adaptive process, allowing individuals to balance situational demands for informativeness and accuracy, and regulate responding to attenuate the effects of reduced memory quality on accuracy (Evans & Fisher, 2011; Fisher, 1996; Weber & Brewer, 2008). From a theoretical perspective, the persistently poor coarse-grained monitoring (together with the lower accuracy of coarse- than fine-grained responses) may reflect a paucity of the information that typically supports the monitoring process (the retrieval of additional contextual detail, retrieval fluency, etc.) and suggest that, depending on situational constraints (e.g., instructions and motivational factors), participants in non-phased regulating conditions may volunteer coarse-grained responses only if the accessible memorial information will not support a more fine-grained response. Further, the varying effects of divided attention on fine- and coarse-grained response accuracy, and the differential confidence–accuracy relations for fine- and coarse-grained responses, reiterate that failing to consider variations in grain size can distort indices of memory and metacognitive performance. From an applied perspective, our findings place an important caveat on the general conclusion that confidence and accuracy are positively related for recall memory responses. Finally, from a methodological standpoint, systematic differences in the results observed for the non-phased and 2-phase procedures in Experiment 2 highlight the need to further examine the extent to which

Table 8

Mixed ANOVA on square root transformed ANDI (Experiment 2).

Effect	df	F	f	p
Between-subjects				
Attention (A)	1	0.71	0.07	.404
Report procedure (RP)	1	20.71	0.38	.000
A × RP	1	0.55	0.06	.462
Error	76	(0.03)		
Within-subjects				
Response granularity (RG)	1	56.68	0.60	.000
RG × A	1	0.55	0.02	.815
RG × RP	1	4.62	0.17	.035
RG × A × RP	1	12.67	0.28	.001
RG error	76	(0.03)		

Note: Values in parentheses represent mean-square errors.

strategic monitoring and control processes – as elucidated by research using the 2-phase paradigm – faithfully characterize responding in more naturalistic reporting contexts. Importantly, the reporting procedure used not only affected the proportion and accuracy of volunteered fine-grained responses, it affected confidence-based discrimination (though not calibration) for these responses once volunteered.

Why might reporting procedure affect participants' responding? By requiring participants to generate both fine- and coarse-grained response alternatives, the 2-phase procedure may alter the nature of the information participants retrieve from memory, and the way they appraise retrieved candidate responses. The satisficing model (Goldsmith et al., 2002; Simon, 1956, 1990) holds that due to participants' preference for informativeness, when answering recall questions fine-grained responses are treated as the “default” response option (and abandoned only if they fail to reach a criterion level of accuracy). Consistent with this model, in the non-phased condition, it appears that when a fine-grained response can be retrieved, participants generally provide it and use confidence to index accuracy. In contrast, participants in the 2-phase condition are required to search for and evaluate both fine- and coarse-grained responses before deciding which to volunteer. Previous research has shown, in various decision-making domains, that directing individuals' attention toward their own cognitive processes, and having participants explicitly consider alternatives, can affect metacognitive assessments and reduce overconfidence in initial responses (Brewer, Keast, & Rishworth, 2002; Carver & Scheier, 1981; Griffin, Dunning, & Ross, 1990; Hoch, 1985). Thus, this externalized process, requiring explicit consideration of the alternative responses, may increase the availability and perceived validity of a coarse-grained response, and decrease confidence in the initial fine-grained response, ultimately increasing the likelihood that the coarse-grained response will be volunteered. Of course, this is a speculative interpretation requiring further investigation.

4.1. Limitations

Four methodological aspects of our research may limit the generalization of the current findings to applied contexts. First, like much previous research, we used relatively simple stimuli and brief retention intervals. We selected these stimuli to provide a balance between experimental control (i.e., clear manipulations of color and quantity) and complexity, but clearly they do not parallel the complexity of genuine witnessed events. Given our short retention intervals (in line with those used by Weber & Brewer, 2008) it is possible that, although we included a manipulation of memory quality, the observed differences between the divided and full attention conditions might fade over time. Similarly, the observed differences between the non-phased and 2-phase conditions might be reduced when retention interval is increased (if non-phased participants become more conservative in their reporting of fine-grained details, see Evans & Fisher, 2011).

Second, although our non-phased regulation paradigm represents a more realistic regulation task than the typical 2-phase procedure, it is more constrained than the level of grain size control afforded in applied settings (i.e., we set boundaries for the coarseness of coarse-grained responses), and in some previous empirical investigations of the issue (e.g., Ackerman & Goldsmith, 2008). Admittedly, this may have reduced the accuracy of coarse-grained responses. However, this approach affords an unambiguous operationalization of participants' regulation of response granularity within the context of this specific task.

Third, we note that our investigation only speaks to grain-size regulation for responses to questions about color and quantity. Although consistent with previous work (e.g., Weber & Brewer, 2008), and relevant for forensically useful details (e.g., the color of a perpetrator's clothing), this represents only a subset of the details about which witnesses may be questioned. Future work might consider grain size regulation for other sorts of details, but it is not obvious how grain size would be operationalized for such details.

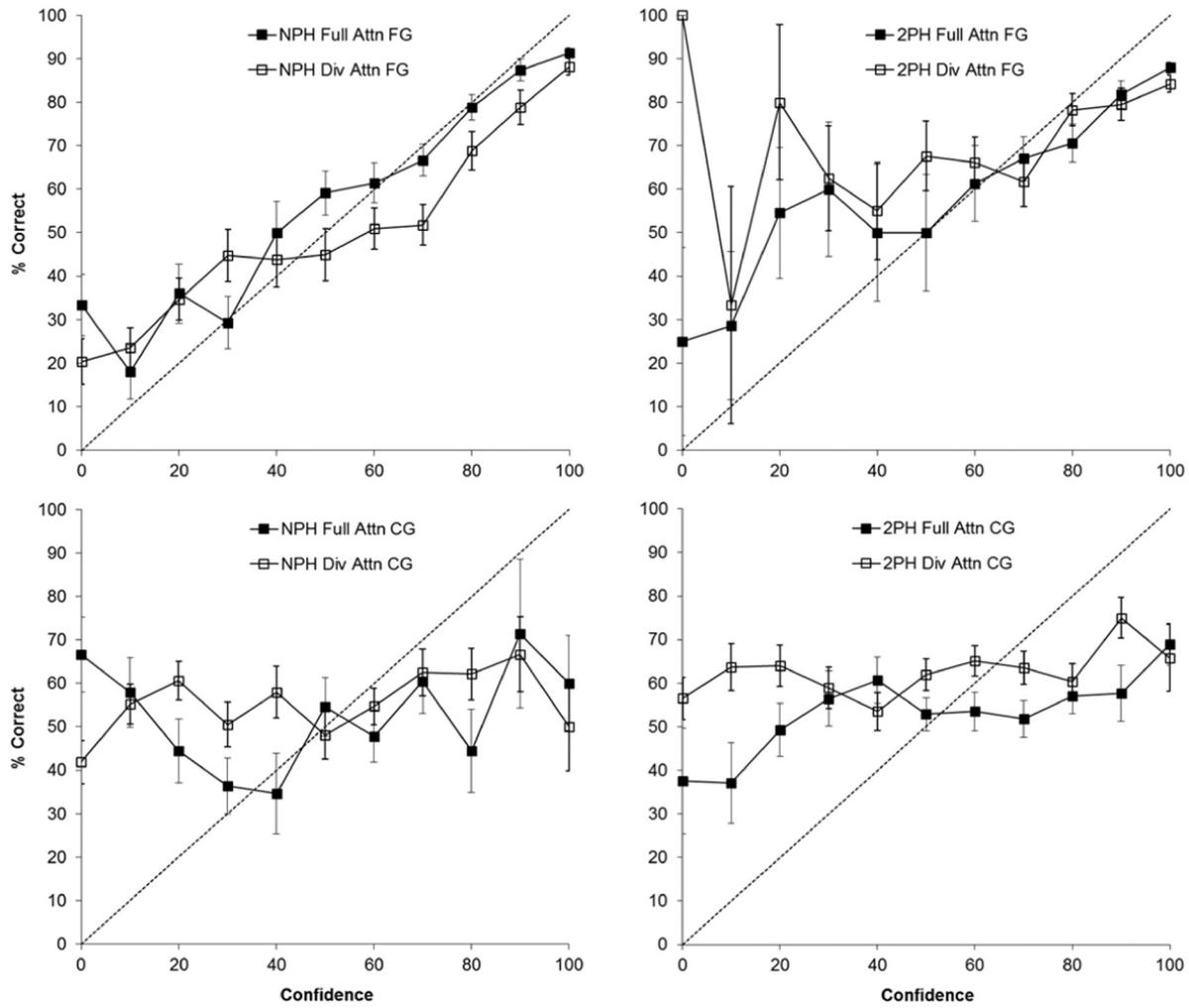


Fig. 5. Calibration curves, based on data collapsed across participants, for volunteered fine-grained (upper panel) and coarse-grained (lower panel) responses in the divided and full attention conditions, for the non-phased (NPH) (left) and 2-phase (2PH) (right) reporting conditions (Experiment 2). Error bars represent standard errors.

Finally, our paradigm did not include an explicit *don't know* response option. The absence of an explicit *don't know* option significantly reduces the likelihood that participants will withhold answers. Thus, our results may not generalize to settings where participants are more conservative in volunteering information. However, like our paradigm, typical cued-recall interviewing practices do not include an explicit *don't know* option, but implicitly allow participants to regulate report option. In fact, investigative interviewers often discourage, implicitly or explicitly, witnesses from providing *don't know* responses (Scoboria & Fisco, 2013). Thus, we argue that this methodological element does not

necessarily undermine the ecological validity of our investigation of grain-size regulation.

4.2. Conclusion

In sum, our findings are encouraging. When memory quality declines, although a willingness to sacrifice accuracy for informativeness increases the chance of fine-grained response errors, *ANDI* and calibration analyses revealed that, in non-phased reporting conditions, witnesses could monitor the accuracy of retrieved information, and that

Table 9
Response frequencies for each confidence level according to response granularity, attention, and reporting procedure (Experiment 2).

Grain size & attention		Confidence level (%)										
		0	10	20	30	40	50	60	70	80	90	100
Self-regulated Fine-grained	Full	45	39	50	58	48	93	109	168	193	167	539
	Divided	59	89	98	67	64	69	112	114	109	104	255
Coarse-grained	Full	30	38	45	55	26	55	69	43	27	7	20
	Divided	98	116	122	95	69	79	139	80	66	30	24
2-Phase Fine-grained	Full	4	7	11	10	10	14	31	85	109	148	666
	Divided	1	3	5	16	20	34	65	73	124	117	318
Coarse-grained	Full	16	27	67	62	84	174	129	143	151	59	90
	Divided	108	80	100	107	129	176	192	159	144	88	38

witness' confidence provided an index of the likely accuracy of volunteered fine-grained responses. Of course, the extent to which effective monitoring results in the volunteering of accurate (cf. inaccurate) responses will depend on the witness' placement of criteria for accuracy and informativeness.

Acknowledgments

James D. Sauer, Department of Psychology, University of Portsmouth, Portsmouth, Hampshire PO12DY, UK; Lorraine Hope, Department of Psychology, University of Portsmouth, Portsmouth, Hampshire PO12DY, UK. James D. Sauer is now at the School of Medicine (Psychology), University of Tasmania, Tasmania, 7001, Australia.

This research was supported by an Economic and Social Research Council Grant RES-000-22-4189. Special thanks to Shannan Greaney for her work on Experiment 2. Thanks also to Erika Näsholm, Laura Burr-Lonnon, and Peter Frizelle for their assistance with data collection.

Correspondence concerning this paper should be addressed to: James D. Sauer, School of Medicine (Psychology), University of Tasmania, Locked Bag 30, Hobart, Tasmania 7001, Australia Email: Jim.Sauer@Utas.edu.au.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.actpsy.2016.05.015>.

References

- Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting – With and without satisfying knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1224–1245. <http://dx.doi.org/10.1037/a0012938>.
- Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The cognitive interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology Crime & Law*, 11, 183–198. <http://dx.doi.org/10.1080/10683160512331329943>.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412. <http://dx.doi.org/10.1016/j.jml.2007.12.005>.
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2013). lme4: Linear mixed-effects models using Eigen and S4 classes. Retrieved from <http://lme4.r-forge.r-project.org>
- Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence–accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied*, 8, 44–56. <http://dx.doi.org/10.1037/1076-898x.8.1.44>.
- Buccione, I., Fadda, L., Serra, L., Caltagirone, C., & Carlesimo, G. A. (2008). Retrograde episodic and semantic memory impairment correlates with side of temporal lobe damage. *Journal of the International Neuropsychological Society*, 14, 1083–1094. <http://dx.doi.org/10.1017/s1355617708080922>.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence–accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, 7, 26–48. <http://dx.doi.org/10.3758/BF03210724>.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self-regulation: A control theory approach to human behavior*. New York: Springer. <http://dx.doi.org/10.1007/978-1-4612-5887-2>.
- Core Team, R. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- 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(3), 501–508. <http://dx.doi.org/10.1002/acp.1722>.
- Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: Evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*, 129, 155–176. <http://dx.doi.org/10.1037/0096-3445.129.2.155>.
- Fisher, R. P. (1996). Implications of output-bound measures for laboratory and field research in memory. *Behavioral and Brain Sciences*, 19, 197. <http://dx.doi.org/10.1017/S0140525X00042229>.
- Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2007). I saw it for longer than you: The relationship between perceived encoding duration and memory conformity. *Acta Psychologica*, 124, 319–331. <http://dx.doi.org/10.1016/j.actpsy.2006.03.009>.
- Gigerenzer, G., Hoffrage, U., & Kleinboelting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506–528. <http://dx.doi.org/10.1037/0033-295X.98.4.506>.
- 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. <http://dx.doi.org/10.1016/j.jml.2005.01.010>.
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: Applied*, 131, 73–95. <http://dx.doi.org/10.1037/0022-3514.59.6.1128>.
- Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classification. *Journal of the American Statistical Association*, 49, 732–764. <http://dx.doi.org/10.2307/2281536>.
- Griffin, D. W., Dunning, D., & Ross, L. (1990). The role of construal processes in overconfident predictions about the self and others. *Journal of Personality and Social Psychology*, 59, 1128–1139. <http://dx.doi.org/10.1037/0022-3514.59.6.1128>.
- Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 719–731. <http://dx.doi.org/10.1037/0278-7393.11.1.4.719>.
- Hope, L., Gabbert, F., & Brewer, N. (2010, November). Interviewing eyewitnesses: Enhancing output quantity and diagnosing accuracy using the grain size technique. Paper presented at the British Psychological Society workshop Metacognition: Where next?. University of Hull.
- Ihlebaek, C., Love, T., Eilertsen, D. E., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory*, 11, 319–327. <http://dx.doi.org/10.1080/09658210244000018>.
- Iidaka, T., Anderson, N. D., Kapur, S., Cabeza, R., & Craik, F. I. M. (2000). The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. *Journal of Cognitive Neuroscience*, 12, 267–280. <http://dx.doi.org/10.1162/089892900562093>.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. <http://dx.doi.org/10.1016/j.jml.2007.11.007>.
- Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford: Oxford University Press, 432.
- Kelley, C. M., & Sahakyan, L. (2003). Memory, monitoring, and control in the attainment of memory accuracy. *Journal of Memory and Language*, 48(4), 704–721. [http://dx.doi.org/10.1016/S0749-596X\(02\)00504-1](http://dx.doi.org/10.1016/S0749-596X(02)00504-1).
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, 103, 490–517. <http://dx.doi.org/10.1037/0033-295X.103.3.490>.
- Koriat, A., Sheffer, L., & Ma'yan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131, 147–162. <http://dx.doi.org/10.1037/0096-3445.131.2.147>.
- Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, 20, 199–212. <http://dx.doi.org/10.1002/acp.1177>.
- Leonesio, R. J., & Nelson, T. O. (1990). Do different metamemory judgments tap the same underlying aspects of memory? *Journal of Experimental Psychology—Learning Memory and Cognition*, 16, 464–470. <http://dx.doi.org/10.1037/0278-7393.16.3.464>.
- Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science*, 9, 215–218. <http://dx.doi.org/10.1111/1467-9280.00041>.
- Luna, K., & Martín-Luengo, B. (2012). Confidence–accuracy calibration with general knowledge and eyewitness memory cued recall questions. *Applied Cognitive Psychology*, 26, 289–295. <http://dx.doi.org/10.1002/acp.1822>.
- Morey, R. D., & Rouder, J. N. (2014). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.9. <http://CRAN.R-project.org/package=BayesFactor>
- Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology*, 20, 973–985. <http://dx.doi.org/10.1002/acp.1263>.
- Pansky, A., Goldsmith, M., Koriat, A., & Pearlman-Avni, S. (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. *European Journal of Cognitive Psychology*, 21, 303–329. <http://dx.doi.org/10.1080/0954144080281183>.
- Perfect, T. J. (2002). When does eyewitness confidence predict performance? In T. J. Perfect, & B. L. Schwartz (Eds.), *Applied Metacognition* (pp. 95–120). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511489976.006>.
- Perfect, T. J., & Stollery, B. (1993). Memory and metamemory performance in older adults: One deficit or two? *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 46, 119–135. <http://dx.doi.org/10.1080/14640749308401069>.
- Perfect, T. J., & Weber, N. (2012). How should witnesses regulate the accuracy of their identification decisions: One step forward, two steps back? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38, 1810–1818. <http://dx.doi.org/10.1037/a0028461>.
- Perfect, T. J., Watson, E., & Wagstaff, G. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology*, 78, 144–147. <http://dx.doi.org/10.1037/0021-9010.78.1.144>.
- Prince, S. E., Tsukiura, T., & Cabeza, R. (2007). Distinguishing the neural correlates of episodic memory encoding and semantic memory retrieval. *Psychological Science*, 18, 144–151. <http://dx.doi.org/10.1111/j.1467-9280.2007.01864.x>.
- Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, 8, 33–43. <http://dx.doi.org/10.1037/1076-898x.8.1.33>.
- Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology*, 81, 587–594. <http://dx.doi.org/10.1037/0021-9010.81.5.587>.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 752–760. <http://dx.doi.org/10.3758/PBR.16.2.752>.

- Scoboria, A., & Fisco, S. (2013). Encouraging and clarifying "Don't Know" responses enhances interview quality. *Journal of Experimental Psychology: Applied*, 19, 72–82. <http://dx.doi.org/10.1037/a0032067>.
- Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review*, 63, 129–138. <http://dx.doi.org/10.1037/h0042769>.
- Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1–19. <http://dx.doi.org/10.1146/annurev.ps.41.020190.000245>.
- Smith, V. L., Ellsworth, P. C., & Kassir, S. M. (1989). Eyewitness accuracy and confidence: Within- versus between-subjects correlations. *Journal of Applied Psychology*, 74, 356–359. <http://dx.doi.org/10.1037/0021-9010.74.2.356>.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson.
- Weber, N., & Brewer, N. (2004). Confidence–accuracy calibration in absolute and relative face recognition judgements. *Journal of Experimental Psychology: Applied*, 10, 156–172. <http://dx.doi.org/10.1037/1076-898X.10.3.156>.
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, 14, 50–60. <http://dx.doi.org/10.1037/1076-898X.14.1.50>.
- Wong, J. T., Cramer, S. J., & Gallo, D. A. (2012). Age-related reduction of the confidence–accuracy relationship in episodic memory: Effects of recollection quality and retrieval monitoring. *Psychology and Aging*, 27, 1053–1065. <http://dx.doi.org/10.1037/a0027686>.
- Yaniv, I., Yates, J. F., & Smith, J. K. (1991). Measures of discrimination skill in probabilistic judgment. *Psychological Bulletin*, 110, 611–617. <http://dx.doi.org/10.1037/0033-2909.110.3.611>.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46, 441–517. <http://dx.doi.org/10.1006/jmla.2002.2864>.