

# Working memory representations persist in the face of unexpected task alterations

Garrett Swan<sup>1</sup> · Brad Wyble<sup>1</sup> · Hui Chen<sup>2</sup>

© The Psychonomic Society, Inc. 2017

**Abstract** It is well known that information can be held in memory while performing other tasks concurrently, such as remembering a color or number during a separate visual search task. However, it is not clear what happens to stored information in the face of unexpected tasks, such as the surprise questions that are often used in experiments related to inattention and change blindness. Does the unpredicted shift in task context cause memory representations to be cleared in anticipation of new information? To answer this question, we ran two experiments where the task unexpectedly switched partway through the experiment with a surprise question. Half of the participants were asked to report the same attribute (Exp. 1 = Identity, Exp. 2 = Color) of a target stimulus in both presurprise and postsurprise trials, while for the other half, the reported attribute switched from identity to color (Exp. 1) or vice versa (Exp. 2). Importantly, all participants had to read an unexpected set of instructions and respond differently on the surprise trial. Accuracy on the surprise trial was higher for the same-attribute groups than the different-attribute groups. Furthermore, there was no difference in reaction time on the surprise trial between the two groups. These results suggest that information participants expected to report can survive an encounter with an unexpected task. The implication is that failures to report information on a surprise trial in many

experiments reflect genuine differences in memory encoding, rather than forgetting or overwriting induced by the surprise question.

**Keywords** Memory: Visual working and short-term memory · Attention: Interactions with Memory · Attention and memory

While the capabilities of working memory have been well studied for cases in which subjects are performing a task that they know in advance (for review, see Baddeley, 2003), there is relatively little data available concerning the reportability of information when participants encounter an immediate surprising question. In other words, when participants are performing a routine task and are asked a surprise question, will the just-stored information be accessible for report? This question is important to answer for several reasons. First, natural experience often forces us to confront surprising situations and it is important to know how reliably information from previous moments can be retrieved after such events. There has been a lot of research on task switches that are expected (see Kiesel et al., 2010, for a review), but not much is known about how working memory responds in the case of a completely unanticipated change in task. Surprising events could potentially cause stored information to be cleared from memory to facilitate the ability to process and store information about the unexpected event. Furthermore, in the cognitive psychology literature, surprise test methodologies are an important tool for explicitly probing memory of stimuli that subjects did not expect to report. Inattention blindness (Mack & Rock, 1998), change blindness (Simons & Levin, 1997), and attribute amnesia (Chen & Wyble, 2015a) have importantly shown the limitations of human visual processing by using surprise tests (e.g., Chen, Swan, & Wyble, 2016; Eitam, Shoval, & Yeshurun, 2015; Eitam, Yeshurun, Hassan, 2013;

---

Paper Submitted to PsyArXiv on the Open Science Framework. Data and methods available at <https://osf.io/7h8qy/>

---

✉ Hui Chen  
chenhui@zju.edu.cn

<sup>1</sup> Department of Psychology, The Pennsylvania State University, State College, PA, USA

<sup>2</sup> Department of Psychology and Behavioral Sciences, Xixi Campus, Zhejiang University, Hang Zhou, People's Republic of China 310007

Shin & Ma, 2016; Swan, Collins, & Wyble, 2016). It is debated whether the inability to answer such surprise questions is due to a failure to encode the information (i.e., a failure of perception; Mack & Rock, 1998) or a loss of the contents of working memory (i.e., amnesia; Jiang, Shupe, Swallow, & Tan, 2016; Wolfe, 1999). The latter possibility arises from the fact that surprise tests effectively impose an unexpected change of task set, in addition to the challenge of reading and comprehending the instructions, either of which might disrupt recent memory. Therefore, determining to what extent information that was clearly stored in memory can be reported in response to a surprise question is crucial for interpreting many psychological phenomena.

Here, an attribute amnesia paradigm (Chen & Wyble, 2016) was altered to test whether or not the contents of working memory are disrupted during a surprise test after subjects have well-developed expectations about what task to perform. Attribute amnesia (AA) reflects a failure to report attributes (e.g., color, letter identity) of an immediately preceding attended stimulus when subjects did not expect to report those attributes (Chen & Wyble 2015a, 2015b, 2016). However, a crucial question that these experiments did not clearly address is whether the memory trace of information that subjects *did* expect to report could persist in response to a surprise question. The current study attempted to answer this question by unexpectedly altering the reporting requirements halfway through an experiment from recognition to recall. An additional manipulation controlled whether the new recall question pertained to information that participants did or did not expect to report.

## Experiment 1

### Method

**Participants** Eighty Pennsylvania State University undergraduates completed Experiment 1 for course credits in accordance with the local Institutional Review Board. Participants were instructed in English and reported normal or corrected-to-normal visual acuity. An additional 19 participants completed the experiment but were removed for having accuracy in the presurprise trials lower than 60% (average performance for excluded participants presurprise accuracy: 34%). The large amount of exclusions for both experiments is likely due to the lack of practice trials.

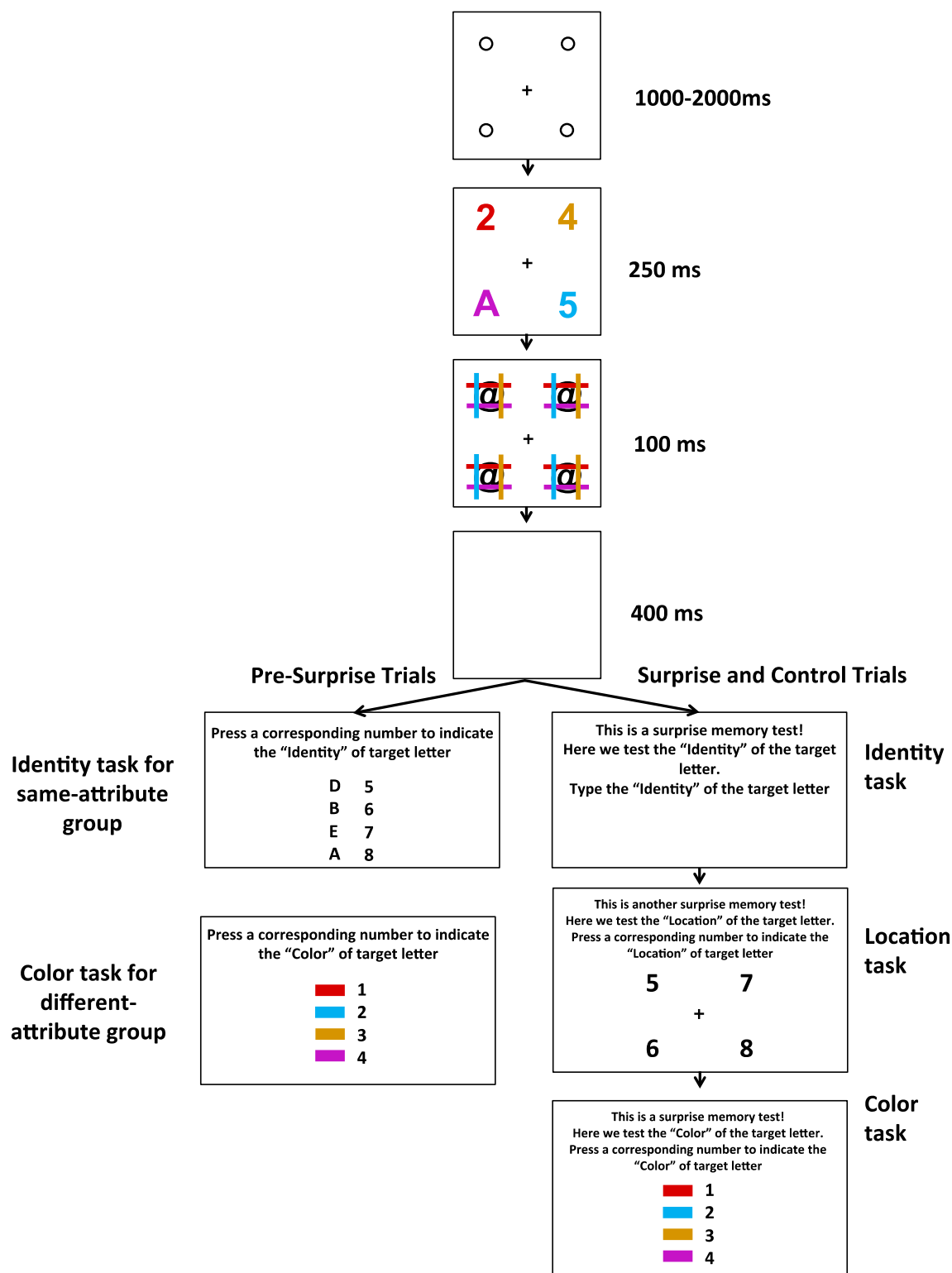
**Apparatus** The stimuli and apparatus follow closely with the physical settings of Chen and Wyble (2015a). Participants were situated 50 cm in a chinrest from a 17-inch cathode ray tube monitor with a resolution of  $1024 \times 768$  at 75 Hz refresh rate. The experiment was presented using MATLAB with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

**Stimuli** In the fixation display, there were four black placeholder circles (radius = 0.31 degrees of visual angle) located at four corners of an invisible square ( $6.25 \times 6.25$ ) and a black central fixation cross (0.62). The stimulus array contained one English letter target (A, B, D, or E;  $0.86 \times 0.62$ ) and three Arabic digit distractors (2–5;  $0.86 \times 0.62$ ). Each stimulus was randomly assigned one of four colors, red (RGB: 200, 0, 0), blue (0, 200, 200), yellow (200, 200, 0), or magenta (190, 45, 200), with no repetition of color on each array. The mask consisted of a black ampersand and a pattern of four colored lines using the above colors, displayed in a hash-mark configuration (see Fig. 1). The location of the letters, digits, and masks were also on the four corners of the invisible square. All stimuli were displayed on a gray background (150, 150, 150).

**Procedure and design** Participants were instructed that a colored letter would appear among three colored digits (see Fig. 1). Each trial began with a fixation display for a duration varied randomly and uniformly between 1,000 and 2,000 ms. Then, the stimulus array appeared for 250 ms, and was replaced by masks for 100 ms. After the masks disappeared, there was a 400 ms retention interval before one or more questions were asked about the array that had just been presented. Participants in same-attribute group were asked to recognize the *identity* of the colored letter by typing in a matching number alongside one of four black letters (see Fig. 1) for the first 11 trials. On trial 12 (i.e., surprise trial), these participants were asked about the same information (letter *identity*) but using recall instead of recognition. Now, participants typed in the letter directly without seeing a list of possibilities (i.e., press the *a* key to report A).

Participants in the different-attribute group were just like participants in the same-attribute group, except that these participants were asked to recognize the *color* of the letter in the 11 presurprise trials by pressing a number alongside one of four colored bars, and then recall its *identity* in the surprise trial using the same question format as for the same-attribute group.

For all participants, the surprise question on trial 12 was also followed by two additional questions. Participants were presented with four digits (1, 2, 3, 4 in randomized order) and asked to report the location of the target letter by typing the corresponding number. Finally, they were asked to recognize the color of the letter. Feedback was presented at the end of each trial informing participants of the correct attribute and response key. For both groups, after the surprise trial, there were four more control trials identical to the surprise trial, thus providing a total of 16 trials. Each group had forty participants. Thus, for all participants, there were 11 presurprise trials, then a surprise trial on Trial 12, and then four control trials that were identical to the surprise trial. Participants were free to use either hand when responding. For all responses,



**Fig. 1** Procedure for Experiment 1. Note the both the same attribute and different attribute had the same sequence of responses on the surprise trial. The only difference between the groups is the attribute reported in

the presurprise trials (i.e., identity for the same-attribute group and color for the different-attribute group). (Color figure online)

except for identity recall, participants were given four possible choices.

We expected that participants would not be able to report the probed attribute on the surprise trial in the different-

attribute group as shown in previous findings of AA, whereas participants in same-attribute group would be able to report the probed attribute because the surprise question now corresponded to the same attribute that was reported in the presurprise trials. Note that all participants received a surprise question on Trial 12 and the format of the surprise question was identical between groups that were asked about the same attribute (*identity*). Therefore any differences in performance must be related to the memory representation formed prior to the surprise question.

## Results

The results of this experiment are depicted in Table 1. The accuracy of the presurprise trials for the same-attribute and different-attribute groups were 91%, and 83%, respectively, which indicates that participants could accurately recognize attributes of the target letter among distractor numbers.

To determine whether reporting the same or different attribute affected accuracy on the surprise trial, a chi-square test was conducted to compare the accuracy on the surprise trial between same-attribute and different-attribute groups. There was a significant effect between these two Groups on the surprise test,  $\chi^2(1, N = 80) = 5.33, p < .05, \phi = .26$ , indicating that the accuracy on the surprise trial was significantly influenced by the identity of response attribute between the presurprise and surprise trials, with the performance being better when they were the same than they were different.

Next, reaction times on the surprise trial were compared between the different groups, where median reaction times across the groups were 9.7 s ( $SEM = .98$ ) and 10.8 s ( $SEM = 1.1$ ) for same-attribute and different-attribute groups, respectively. In an independent one-tail  $t$  test, there was no significant effect of presurprise attribute,  $t(78, N = 80) = .68, p = .25, \phi = .09$ . Thus, the effects on accuracy do not seem to be related to differences in reaction time.

**Table 1** Results of Experiment 1

Same-attribute group of Experiment 1						
	Presurprise	Surprise	Contr. 1	Contr. 2	Contr. 3	Contr. 4
Location	N/A	40% <sup>2</sup>	83%	90%	90%	95%
Color	N/A	35% <sup>3</sup>	85%	85%	75%	83%
Identity	91%	75% <sup>1</sup>	85%	95%	90%	88%
Different-attribute group of Experiment 1						
	Presurprise	Surprise	Contr. 1	Contr. 2	Contr. 3	Contr. 4
Location	N/A	78% <sup>2</sup>	93%	88%	93%	98%
Color	83%	60% <sup>3</sup>	83%	80%	90%	88%
Identity	N/A	50% <sup>1</sup>	80%	75%	93%	83%

*Note.* Superscripts indicate the order that questions were asked in the surprise and all following control trials. Contrs. 1–4 correspond to the four control trials

In addition to comparing the surprise trial accuracy across groups, accuracy on the surprise trial was compared to the accuracy on the first postsurprise trial to determine whether AA was replicated. In the different attribute group, 20 out of 40 participants (50%) accurately selected the color, whereas 32 out of 40 (80%) were correct on the first postsurprise trial. In a chi-square test, this difference is significant,  $\chi^2(1, N = 80) = 7.91, p < .01, \phi = .31$ , replicating the AA phenomenon (Chen & Wyble, 2015a). In the same-attribute group, 30 out of 40 (75%) were correct, whereas 34 out of 40 (85%) were correct on the first postsurprise trial and this difference was not significant,  $\chi^2(1, N = 80) = 1.25, p = .26, \phi = .13$ .

## Experiment 2

The results of Experiment 1 show that memory traces can persist in the face of unexpected immediate surprise questions if participant expected to report that information. However, this could be the result of using the identity as the report attribute. Experiment 2 was designed to replicate Experiment 1 but using the color of the target letter as the report attribute on the surprise trial for both groups.

## Method

The methods for Experiment 2 are exactly like Experiment 1, except where follows. Eighty new Pennsylvania State University undergraduates completed Experiment 2 for course credits. An additional 36 participants completed the experiment but were removed for having accuracy in the presurprise trials lower than 60% (average excluded participants presurprise accuracy: 31%) and replaced.

Participants in same-attribute group in Experiment 2 were asked to recognize the *color* of the colored letter by typing in a matching number alongside one of four colored bars (see Fig. 1) for the first 11 trials. On Trial 12, these participants were asked about the same information (letter *color*) but using recall instead of recognition. Now, participants typed in the first letter of the color word (i.e., press the *r* key if the target letter was colored red).

Participants in the different-attribute group were just like participants in the same-attribute group except that these participants were asked to recognize the *identity* of the letter in the 11 presurprise trials by pressing a number alongside one of four black letters, and then switched to recall its *color* in the surprise trial by typing in the first letter of the color word as in same-attribute group.

## Results

The results of Experiment 2 are depicted in Table 2. The accuracy of the presurprise trials for the same-attribute and

**Table 2** Results of Experiment 2

Same-attribute group of Experiment 2						
	Presurprise	Surprise	Contr. 1	Contr. 2	Contr. 3	Contr. 4
Location	N/A	43% <sup>2</sup>	68%	80%	78%	90%
Color	83%	68% <sup>1</sup>	80%	85%	75%	83%
Identity	N/A	45% <sup>3</sup>	55%	73%	63%	73%
Different-attribute group of Experiment 2						
	Presurprise	Surprise	Contr. 1	Contr. 2	Contr. 3	Contr. 4
Location	N/A	63% <sup>2</sup>	78%	73%	88%	85%
Color	N/A	43% <sup>1</sup>	65%	73%	88%	90%
Identity	85%	53% <sup>3</sup>	70%	78%	75%	78%

Note. Superscripts indicate the order that questions were asked in the surprise and all following control trials. Contrs. 1–4 correspond to the four control trials

different attribute groups were, 83%, and 85% respectively, which indicates that participants could accurately recognize attributes of the target letter among distractor numbers. The same sets of analyses used in Experiment 1 were also used for Experiment 2. All of the effects found in Experiment 1 were replicated in Experiment 2.

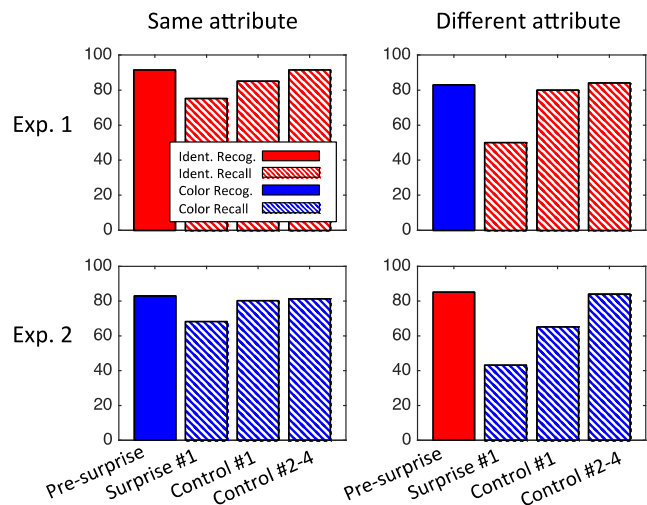
When comparing accuracy on the surprise trial across groups, there was a significant effect,  $\chi^2(1, N = 80) = 5.05$ ,  $p < .05$ ,  $\phi = .25$ . Participants were more accurate on the surprise trial (27 out of 40 were correct) if the attribute was the same between presurprise and surprise trial compared to the different-attribute group (17 out of 40 were correct), which is consistent with the findings of Experiment 1.

There was no effect on reaction time between the same-attribute (median = 10.5 s;  $SEM = 1.1$ ) and different-attribute groups (median = 12.7;  $SEM = 1.2$ ),  $t(78, N = 80) = 1.26$ ,  $p = .11$ , which replicates the results from Experiment 1.

Next, accuracy on the surprise trial was compared to the accuracy on the first postsurprise trial to determine whether AA was replicated. In the different attribute group, 17 out of 40 (43%) were correct, whereas 26 out of 40 (65%) were correct on the first postsurprise trial, which was significant,  $\chi^2(1, N = 80) = 4.07$ ,  $p < .05$ ,  $\phi = .23$ , in a chi-square test replicating AA. In the same-attribute group, 27 out of 40 (68%) were correct, whereas 32 out of 40 (80%) were correct on the first postsurprise trial, which was not significant,  $\chi^2(1, N = 80) = 1.61$ ,  $p = .2$ ,  $\phi = .14$ , in a chi-square test. These results thus mirror the results from Experiment 1.

### Combined analysis

To determine if there was a difference between the pattern of Experiments 1 and 2 (see Fig. 2), a log-linear analysis was used. Log linear analyses can test for interactions in categorical data with multiple dimensions (Goodman, 1979; Pomer, 1984; White, Pesner, & Reitz, 1983). In a log-linear analysis of the 2 (identity of response attribute between the



**Fig. 2** Accuracy between groups for different Experiments. *Presurprise* refers to the average presurprise accuracy and control #2-4 refers to the average control Trials 2–4. The largest standard error in the presurprise trials is less than 2%. (Color figure online)

presurprise and surprise trials: same, different)  $\times$  2 (response: correct, incorrect)  $\times$  2 (Experiment: 1, 2) contingency table (see Table 3) was conducted. There was a significant effect between the identity of the response attribute and the response ( $G^2 = 11.5$ ,  $p < .005$ ), replicating the individual chi-squares tests for each experiment. However, there was no significant effect of experiment and response ( $G^2 = .9$ ,  $p = .33$ ), suggesting that there was no difference in overall performance on the surprise trial as a function of the presurprise attribute.

### General discussion

Two experiments were conducted to explore whether or not a visual memory can be retrieved in response to a surprise test wherein participants need to read a new set of instructions and respond in a new way (see Table 4). Results from the same-attribute groups showed that participants could accurately report the attribute they expected to report, despite encountering a surprise test. On the other hand, the different-attribute groups had difficulty responding to the unexpected change in the report task, presumably because they were now being asked about an attribute they had not expected to report, which replicates the AA effect (Chen & Wyble, 2015a, 2016).

**Table 3** Surprise trial across experiments

	Same attribute		Different attribute	
	Correct	Incorrect	Correct	Incorrect
Exp. 1	30	10	20	20
Exp. 2	27	13	17	23



**Table 4** Summary of the results

Reported presurprise	Reported attribute in the surprise test	
attribute	Identity	Color
Identity	AA was eliminated (Exp. 1)	Replicated AA (Exp. 2)
Color	Replicated AA (Exp. 1)	AA was eliminated (Exp. 2)

*Note.* AA corresponds to attribute amnesia

These results are important for multiple reasons. First, these results show that information that is already stored in working memory can still be reported despite an unexpected change in the response task. Participants that reported the same attribute in the presurprise and surprise trials were significantly more accurate on the surprise trial than the different attribute group, despite the unexpected shift for both groups from recognition to recall and the need to use different response keys. Whereas it could have been the case that the surprise trial effectively cleared the contents of working memory, since participants had to perform several complex tasks, including reading of the surprise question and planning a new response using a different set of keys, we found instead that the information content participants had planned to report was still available. This suggests that the use of surprise questions is not the primary reason for an inability to report information in inattentional blindness, change blindness, and attribute amnesia experiments. Instead, these results support the idea that the primary factor in determining whether or not information is reportable in a surprise test is how congruent that attribute is with our task set (Most et al., 2001) and the strength of its consolidation (Chen & Wyble, 2016).

The result also highlights the robustness of working memory representations in the face of unexpected processing demands, which is rarely explored in task switching. Typically in task switching experiments, costs in performance are observed during expected switches from one task to another. Even in cases where the switches occur at unexpected times, there is still an expectation that the switch might occur. However here, switches in the type of response (e.g., from recognition to recall) were completely unexpected and yet the information was still available for report.

This finding also supports a fractionation of working memory into specific memory stores (Baddeley, 2003). According to these models, reading and visuospatial information are processed in separate stores, so there should be minimal interference when reading the unexpected task instruction and maintaining the attribute. This is contrary to theories of working memory that have only a single store (e.g., Cowan, 2001), since such models would have predicted more dramatic interference on the surprise trial even for participants in the same-attribute group. However, it is possible that the methodology used here is not sensitive enough to capture the interference and that more

precise measures of retrieval, such as delayed estimation (Wilken & Ma, 2004), may be able to measure any costs.

Furthermore, these results clearly indicate that memories held in working memory are content addressable, which means that information can be retrieved by its content and is not just stored as a pre-mapped response (e.g., Swan & Wyble, 2014). Here, participants that reported the same attribute in the pre-surprise and surprise trials were able to retrieve that information using recall instead of recognition. Content addressability is an important constraint when considering the plausibility of models of working memory because it means that the information is stored in a generic format that can be accessed in several ways.

Last, there was no significant interaction of reaction time on the surprise trial between the different groups. This result suggests that the inability to report an attribute on the surprise trial for participants that reported a different attribute in the pre-surprise trials was not a function of the amount of time to respond (cf. Jiang et al., 2016). Instead, these results suggest that the critical component to be able to report an attribute in a surprise test is the consolidation of that attribute into working memory (Chen & Wyble, 2016), given that information which participants expected to report typically survived a substantial unexpected delay during which a surprise question had to be processed.

## Conclusion

The experiment presented here demonstrated that a surprise test, which often involves reading a new set of instructions and responding in a new way, is not likely to be the primary cause of a failure to report information in a surprise test. This result increases our confidence in the surprise method as a means to probe the contents of memory for immediately preceding stimuli and also suggests that memory representations are robust to unexpected interference and can be accessed in unexpected ways.

**Acknowledgements** This project was supported by National Science Foundation Grant 1331073. We gratefully acknowledge the Open Science Framework for hosting the manuscript and data files during publication. The project, including the data and all scripts for the experiment can be found here: [https://osf.io/aujzg/?view\\_only=f32516d5dbbf438fb6e223f181cdaa76](https://osf.io/aujzg/?view_only=f32516d5dbbf438fb6e223f181cdaa76)

**Author contributions** All authors developed the study concept, contributed to the writing and editing, and interpretations of the data. Hui Chen and Garrett Swan performed the programming, data collection, and analysis of the data. All authors approved the final version of the manuscript for submission.

## Compliance with ethical standards

**Conflicts of interest** The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

## References

- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, 4(10), 829–839.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10, 433–436.
- Chen, H., Swan, G., & Wyble, B. (2016). Prolonged focal attention without binding: Tracking a ball for half a minute without remembering its color. *Cognition*, 147, 144–148.
- Chen, H., & Wyble, B. (2015a). Amnesia for object attributes failure to report attended information that had just reached conscious awareness. *Psychological Science*, 26(2), 203–210. doi:10.1177/095679761456064
- Chen, H., & Wyble, B. (2015b). The location but not the attributes of visual cues are automatically encoded into working memory. *Vision Research*, 107, 76–85.
- Chen, H., & Wyble, B. (2016). Attribute amnesia reflects a lack of memory consolidation for attended information. *Journal of Experimental Psychology. Human Perception and Performance*, 42, 225–234.
- Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Science*, 24(1), 87–114.
- Eitam, B., Shoval, R., & Yeshurun, Y. (2015). Seeing without knowing: task relevance dissociates between visual awareness and recognition. *Annals of the New York Academy of Sciences*, 1339(1), 125–137.
- Eitam, B., Yeshurun, Y., & Hassan, K. (2013). Blinded by irrelevance: Pure irrelevance induced “blindness”. *Journal of Experimental Psychology. Human Perception and Performance*, 39(3), 611.
- Goodman, L. A. (1979). Simple models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74(367), 537–552.
- Jiang, Y. V., Shupe, J. M., Swallow, K. M., & Tan, D. H. (2016). Memory for recently accessed visual attributes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(8). doi:10.1037/xlm0000231
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching—A review. *Psychological Bulletin*, 136(5), 849–874.
- Mack, A., & Rock, I. (1998). *Inattention blindness* (Vol. 33). Cambridge, MA: MIT Press.
- Most, S. B., Simons, D. J., Scholl, B. J., Jimenez, R., Clifford, E., & Chabris, C. F. (2001). How not to be seen: The contribution of similarity and selective ignoring to sustained inattention blindness. *Psychological Science*, 12(1), 9–17.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Pomer, M. I. (1984). Demystifying loglinear analysis four ways to assess interaction in a  $2 \times 2 \times 2$  table. *Sociological Perspectives*, 27(1), 111–135.
- Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, 16(5), 10.
- Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, 1(7), 261–267.
- Swan, G., & Wyble, B. (2014). The binding pool: A model of shared neural resources for distinct items in visual working memory. *Attention, Perception, & Psychophysics*, 76(7), 2136–2157.
- White, D. R., Pesner, R., & Reitz, K. P. (1983). An exact significance test for three-way interaction effects. *Cross-Cultural Research*, 18(2), 103–122.
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 11.
- Wolfe, J. M. (1999). Inattention blindness. In V. Coltheart (Ed.), *Fleeting memories* (pp. 71–94). Cambridge, MA: MIT Press.