rijksuniversiteit
groningen

What is Forgotten in Attribute Amnesia

B.E.L. Koot

Master Thesis – Applied Cognitive Neuroscience

S3125475
July 2022
Department of Psychology
University of Groningen
Examiner/Daily supervisor:
Joost de Jong, MSc.

Abstract

The phenomenon of attribute amnesia suggests that our memory for attended objects might not be as precise as we intuitively believe. Although not all information appears to be lost in attribute amnesia, the way this residual information for stimulus identity is represented in memory requires further scrutinizing. One possibility is that an internal statistical model containing an average representation of the attended identities is automatically created over time. As such, participants would still exhibit attribute amnesia for stimulus identity, but their incorrect responses would be biased toward the previously attended letters. To test this, we replicated Chen and Wyble's (2015) attribute amnesia paradigm, but only showed two target letters throughout the experiment. Additionally, when probing the stimulus identity, we included two letters that had not been shown throughout the experiment. The results indicated that participants were indeed poor at reporting the exact letter identity when there was no expectation to do so. However, there seemed to be a bias toward the previously attended letters. Crucially, this bias reflected their hypothesized internal distribution. Taken together, our results suggest that participants indeed automatically store the attended identities in a representation that reflects an internal statistical model.

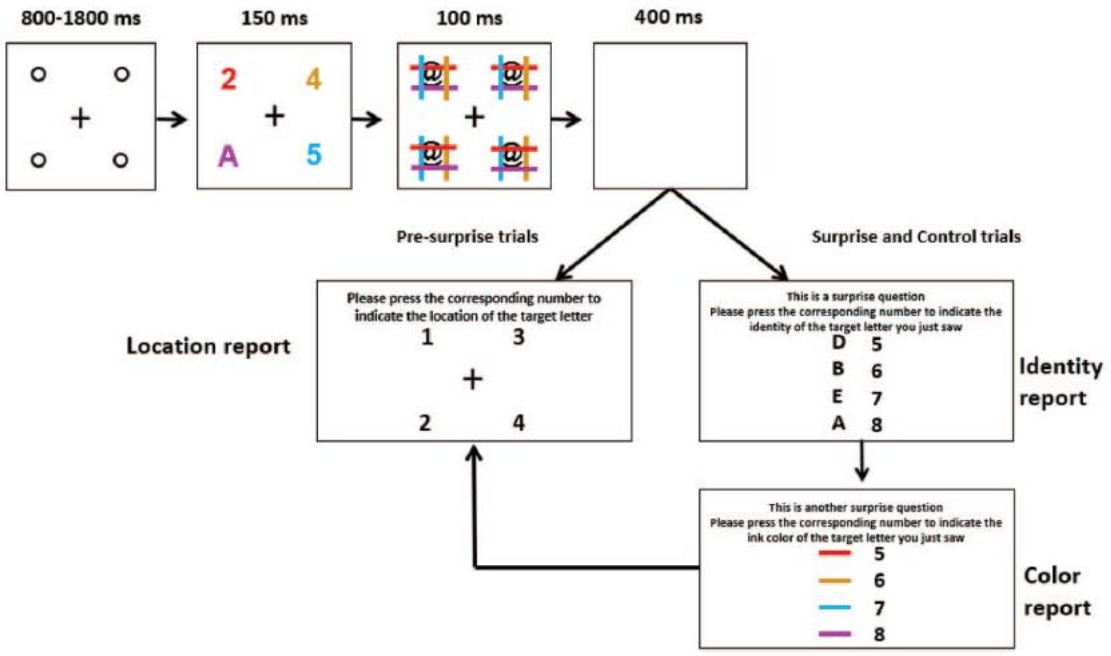
What is Forgotten in Attribute Amnesia

For a long time, it was commonly accepted that attended information reaches consciousness and is subsequently stored in working memory (Wang et al., 2021). The results from Chen and Wyble (2015) challenge this assumption by demonstrating that having attended to certain information does not ensure it is reportable. Chen and Wyble (2015) named this phenomenon ‘’attribute amnesia’’ and their findings demonstrate the discrepancy between what we see and evidently become aware of, and what we can report. Importantly, similar findings have been obtained when attending to faces, which could bear real-world consequences in situations where an eyewitness is supposed to identify the perpetrator from a lineup (Tam et al., 2021). Therefore, gaining more insight into what is and is not remembered in attribute amnesia appears to be of crucial importance.

In a series of experiments, Chen and Wyble (2015) found that participants were surprisingly poor at reporting a salient stimulus attribute when they did not expect to do so. During one such experiment, participants were briefly shown a stimulus array containing three numbers and one target letter, all in a different color. A schematic overview of this experiment is depicted in Figure 1. Participants were instructed to remember the location of the letter and were subsequently required to indicate its location on the screen by pressing the corresponding key. This continued for 155 trials, after which the participants unexpectedly received two multiple-choice questions regarding the identity and color of the target letter they had just seen. Most participants were unable to report the task-irrelevant stimulus attribute (color). Remarkably and unexpectedly, the performance for the target-defining stimulus attribute (identity) was found to be very poor as well. Considering the fact that participants were highly accurate (96%) in reporting the location of the letter, Chen and Wyble (2015) argued that the letter had reached conscious awareness, and thus rejected the possibility that participants had not seen the letter. Crucially, after the surprise trial, participants received the same two questions over four trials as a control. Compared to the surprise trial, participants' ability to report the target letter’s identity and color increased dramatically, suggesting that the inability to correctly answer the surprise question is not due to working memory being overloaded (Chen & Wyble, 2016).

Figure 1

A Schematic Overview of Chen and Wyble's (2015) Attribute Amnesia Paradigm


800-1800 ms
150 ms
100 ms
400 ms
2
4
A
5
@
Pre-surprise trials
Surprise and Control trials
Location report
Please press the corresponding number to indicate the location of the target letter
1
3
2
4
This is a surprise question
Please press the corresponding number to indicate the identity of the target letter you just saw
D 5
B 6
E 7
A 8
Identity report
This is another surprise question
Please press the corresponding number to indicate the ink color of the target letter you just saw
5
6
7
8
Color report

Note. From *Expecting the Unexpected: Violation of Expectation Shifts Strategies Toward Information Exploration* by, Chen et al. (2019b).

The question then remains how it is that most participants were unable to report the target-defining attribute of a stimulus they had just observed. In an attempt to explain this counterintuitive finding, Chen and Wyble (2016) hypothesized that information that is irrelevant for the task at hand, does not enter working memory. Although the identity of the target was the target-defining attribute in Chen and Wyble's (2015) paradigm, it was not necessary to remember this attribute to perform well on the task. Instead, it was sufficient to only remember the location of the target. In fact, because the memory for the location of the target seems to survive when presented with an unexpected question, Chen and Wyble (2016) proposed that encoding stimulus location is obligatory. In other words, the location of a salient stimulus is encoded automatically. Due to participants' poor ability to report the identity and color of the target in a surprise question, Chen and Wyble (2016) concluded that such automatic encoding does not occur for stimulus attributes. However, recent findings suggest otherwise.

Implicit Memory for Stimulus Attributes

Contrasting Chen and Wyble (2016), Jiang et al. (2016) demonstrated that participants appear to retain some information about the identity of the target, irrespective of its expected relevancy. In a replication of Chen and Wyble's (2015) paradigm, Jiang et al. (2016) found that when the identity of the target was repeated over consecutive trials, participants were quicker at reporting its location. The presence of this repetition priming effect suggests that at least some memory for the attended stimulus was retained. However, this implicit residual information might be encoded in such a weak state, Jiang et al. (2016) argued, that the surprise of the unexpected question causes this implicit memory trace to be erased from working memory, which would explain the poor performance on the surprise trial. Nevertheless, subsequent studies have shown that the representation of the attended stimulus is strong enough to survive the impact from a surprise question (Swan et al., 2017; Chen et al., 2019b). In fact, whereas Jiang et al. (2016) hypothesized that the implicit memory trace for stimulus information is so fragile that it is overwritten by new information, Harrison et al. (2021) found that this residual information is encoded in such a way that it can influence behavior.

In an attempt to shed more light on the nature of the memory representation for attended information, Harrison et al. (2021) extended Chen and Wyble's (2015) paradigm by including an additional search task. In the first task, participants were required to remember the location of the colored letter among three other letters. Immediately after this and thus before the location of the target was probed, an additional search task appeared. This time, participants needed to spot the target shape among five distractors and indicate the orientation of a line within the target as quickly as possible. Finally, the location of the letter from this first task was probed. Importantly, the shapes in the second task were all colored such that either the color of the target shape, a distractor shape, or no shape at all matched the color of the target letter. After 180 trials, a surprise question asked participants to indicate the color of the letter. While their performance on the surprise task was indicative of attribute amnesia, the color of the target letter was not immediately forgotten. Rather, Harrison et al. (2021) found that when the color of the target letter matched the color of the target shape, the reaction times were faster, whereas the opposite was true when the color matched a distractor. Therefore, Harrison et al. (2021) extended the initial findings from Jiang et al. (2016) by pointing out that residual information for the attended stimulus attributes is encoded into working memory such that it is able to create a bias toward the previously attended attributes.

Explicit Memory for Spatial Information

The results from Jiang et al. (2016) and Harrison et al. (2021) imply that not all information is lost in attribute amnesia. In fact, under certain conditions, even completely irrelevant stimulus information such as the color of the distractors seems to leave an implicit memory trace (Harrison et al., 2021; Shin & Ma, 2016; Born et al., 2020; Swan et al., 2016). Nevertheless, the previously discussed literature relied on implicit measures of memory by examining reaction times. Therefore, it remains unclear whether this implicit information is explicitly retrievable.

O'Donnell et al. (2021) shed some light on this question by examining whether or not the display configuration of a trial is automatically encoded in a reportable memory trace. In order to achieve this, they used Chen and Wyble's (2015) attribute amnesia paradigm, where participants needed to report the location of targets while the layout of the screen switched randomly between two predetermined configurations. After doing so for a number of trials, participants received a surprise question, asking them to indicate the identity of the target and more importantly, the layout of the display they had just seen. In one of their experiments, O'Donnell et al. (2021, Experiment 3) included two novel answer options when probing the display configuration, such that the answers included the two configurations participants had seen throughout the experiment, and two they had never seen before. Although attribute amnesia for the display configuration was found to be present, most participants exhibited some knowledge of what they had seen. More specifically, a significant number of participants who incorrectly reported the display configuration, mistakenly chose the other display that was shown throughout the experiment instead. In other words, despite the fact that they were unable to report the exact configuration, they were able to make a distinction between what they had and had not seen. This suggests that spatial information such as the trial layout is automatically encoded, but not specific to a particular trial (O'Donnell et al., 2021).

The finding that overall residual memory can be reported, but not for a specific trial, has also been demonstrated by Chen and Howe (2017). In an attempt to further examine the effect of familiarity on attribute amnesia as described by Chen and Wyble (2016), Chen and Howe (2017) used a novel target on each trial. Familiarity, in this case, refers therefore to the repeated exposure to certain stimuli, such that they become familiar. Whereas Chen and Wyble (2016) still found an attribute amnesia effect for letters, despite eliminating the influence of familiarity, Chen and Howe (2017) found the opposite for images of animals. Though, it should be mentioned that this discrepancy might be due to the difference in perceptual processing of simple and complex stimuli (Chen et al., 2019a), their experiments differed in one fundamental aspect. While Chen and Wyble (2016) included the previously presented letters as distractors when probing the identity on the surprise trial, Chen and Howe (2017) used distractors that had not been shown throughout the experiment. Based on both their own results and those from Chen and Wyble (2016), Chen and Howe (2017) therefore hypothesized that attribute amnesia reflects the inability to make a distinction between familiar stimuli for a specific trial. More specifically, Howe and Chen (2017) concluded that participants are very aware of what they have seen, they are simply unable to remember the exact moment. From a theoretical perspective, they argued that a long-term memory trace was automatically formed for each attended image (Oberauer, 2002). However, this memory trace would only be bound to a specific trial if it was necessary for the task at hand. This expectancy-based binding hypothesis (Chen et al., 2016) could explain the non-specific residual memory for display configurations found by O’Donnell et al. (2021).

Based on these results, it appears that under certain conditions, residual information is encoded in such a way that its memory trace is explicitly retrievable. Nevertheless, in an experiment similar to the novelty manipulation by Chen and Wyble (2016), Chen et al. (2019a, Experiment 2a) still obtained an attribute amnesia effect, despite using three novel answer options, besides the correct option, on the surprise trial. This suggests that the process that enables the encoding of attended information into a summary format needs further elaboration.

Statistical Learning

In research regarding human cognition and behavior, it has been found that we automatically extract statistical information from regularities in our environment (Sherman et al., 2020). This statistical learning leads to the gradual and implicit formation of an internal model that contains a compressed representation of the learned regularities (Brady et al., 2009). As such, statistical learning allows for an efficient representation of the attended information, which might be an efficient manner to organize cognitive resources. The main consequence of obtaining such an internal model is the ability to make predictions about upcoming events (Sherman et al., 2020). In fact, statistical learning could facilitate the perceptual processing of upcoming stimuli if they are consistent with the obtained internal model (de Lange et al., 2018). That is, stimuli that have appeared regularly throughout the experiment are processed with more ease in future trials. Conversely, when a novel stimulus is presented after having established an internal model, responses are often biased toward the prior attended information within that internal model (de Lange et al., 2018). Taken together, it is possible that the summary format of attended display configurations found by O’Donnell et al. (2021, Experiment 3) reflects the automatic formation of an internal model via statistical learning. Furthermore, this internal model might then facilitate participants in determining what they have and have not seen, but might be insufficient for making a distinction between information within the model itself, which ultimately leads to attribute amnesia.

Statistical learning could also explain why Chen et al. (2019a, Experiment 2a) still obtained an attribute amnesia effect, despite including three novel distractors as answer options when probing stimulus identity. Namely, due to the fact that Chen et al. (2019a, Experiment 2a) presented a novel target letter on each trial and thus eliminated regularities, it is possible that participants were unable to construct a reliable internal model and thus were unable to make a distinction between familiar and novel stimuli. Moreover, throughout their experiment Chen et al. (2019, Experiment 2a) showed 11 possible target letters. Even if those letters were all equally repeated across trials, attribute amnesia might still be evident, considering the fact that an internal model might be probabilistic of nature (de Lange et al., 2018). As such, the theoretical probability assigned to each letter within such an internal model might be so small, that it offers no help in distinguishing between the distractors on the surprise trial. Therefore, while O’Donnell et al. (2021) provided some evidence for an internal statistical model for display configurations, the results obtained by Chen et al. (2019a) still leave the question whether the same occurs for stimulus identity.

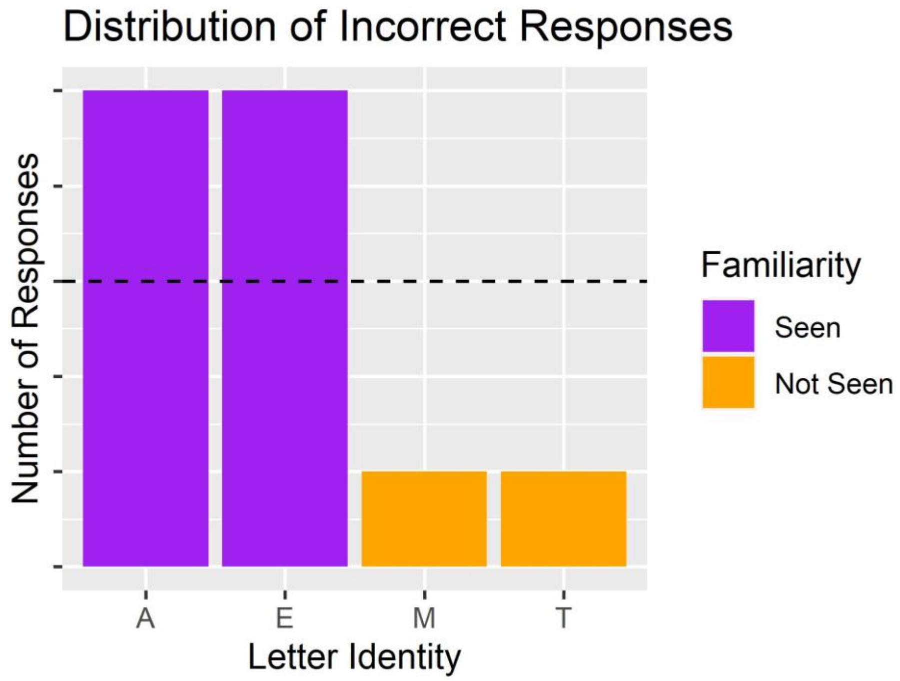
The Current Study

The current study aimed to find out whether the identity of a stimulus is encoded in an explicitly retrievable form that represents the internal statistical distribution of the attended identities. To achieve this, we replicated the second experiment from Chen and Wyble (2015). Over a course of 32 trials, participants were required to remember the location of the target letter, among three digits, which all differed in color. Then, on the 33rd trial, they were unexpectedly asked to indicate the identity and color of the letter. However, the current experiment differed from Chen and Wyble (2015) on two crucial aspects. First, in order to obtain an internal distribution of letter identities that could be used for report, we only showed two possible target letters throughout the experiment. Secondly, to test the existence of such a distribution, we included two novel letters, besides the previously presented letters, as foil answer options when probing the identity.

Because the memory trace containing the distribution would not be specific to a particular trial, we still expected to find attribute amnesia for both the target-defining attribute (identity) and the irrelevant attribute (color). However, though this memory trace might prevent participants from making a distinction between prior attended identities, we hypothesized that it would facilitate them in distinguishing between familiar and novel letters. As such, if stimulus identity is automatically extracted and stored in an internal model of attended information, we expected the distribution of errors in reporting the identity to be biased toward the familiar letters (Figure 2). If no such internal model is created, then we expect the distribution of errors to be uniformly distributed around chance-performance.

Figure 2

The Hypothesized Biased Distribution of Incorrect Answers for The Identity-Task



Note. This figure represents a visualization of the hypothesis that participants are able to distinguish between prior attended and novel letters. The data used to obtain the figure is fictional. The dashed black line represents the uniform distribution based on chance-performance (25%).

Method

Participants

In total, 19 participants were first-year psychology students from the University of Groningen who participated for partial course credit. Three participants were obtained by means of a convenience sample, making the total sample size 22 participants. All participants had normal or corrected-to-normal vision and were naïve to the purpose of the experiment. Ethic approval was obtained by the Ethics Committee of the Faculty of Behavioural and Social Sciences at the University of Groningen, before the participants were recruited.

Apparatus

The stimuli were presented on a 27 inch LCD computer screen, with a resolution of 1024 x 768 pixels. Participants were seated at approximately 60 cm from the screen and used a computer keyboard to indicate their responses. The experiment was programmed in OpenSesame (Version 3.3.10; Mathôt et al., 2012).

Stimuli

At the start of each trial, four placeholders would appear around a fixation dot (angular size of 0.25°) along the four corners of an invisible square (6.25° × 6.25°) on a medium-gray background. The subsequent stimulus screen contained three Arabic numbers randomly selected from a set of four numbers (‘2’, ‘3’, ‘4’ and ‘5’), which functioned as the distractors, and one target letter randomly selected from a set of two letters (‘A’ and ‘E’). The angular size of all stimuli was approximately 21°. Additionally, at the start of each trial the color assigned to the stimuli was randomly chosen from a predetermined set (red, blue, yellow or magenta).

Procedure

Before initiating the experiment, participants were provided with an information sheet that explained the nature and purpose of the study. However, in order to maintain the element of surprise on which the current study relies heavily, we intentionally omitted the fact that our actual purpose is to quantify what is remembered in attribute amnesia. Instead, participants were told the experiment measured how well they remember briefly presented information. After having read the information sheet, they could indicate their willingness to participate by signing the informed consent sheet. The researcher who was present at the location then verbally instructed the participants in either Dutch or English about the to-be-performed task.

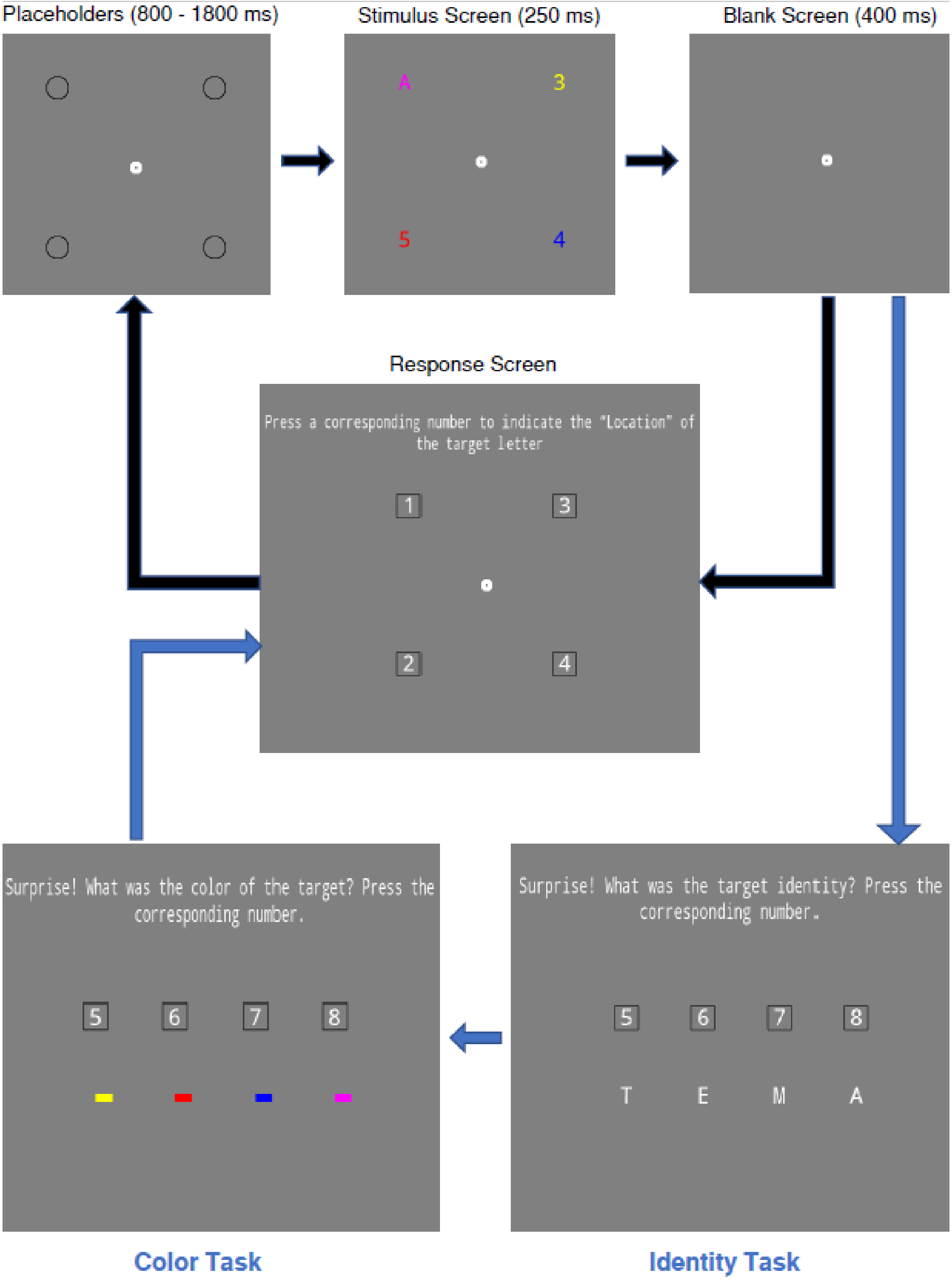
The details of the current study followed, to a large extent, the study conducted by Chen and Wyble (2015). The suggestions by Tam et al. (2021) to reduce the number of words on the surprise trial and to present the answer options horizontally, rather than vertically were also incorporated. A schematic overview of the current experiment is depicted in Figure 3. First, in order to prevent participants from timing the onset of the stimulus screen, and therefore encourage them to remain focused, the placeholders were presented with a duration varying between 800 and 1800 milliseconds. Subsequently, participants would see a screen containing one letter and three digits, which were all assigned a different color. During this brief exposure (250 milliseconds), participants needed to remember the location of the letter and indicate its location by pressing the key that corresponded to the location on the screen (‘1’, ‘2’, ‘3’ or ‘4’). An inter-stimulus-interval (400 milliseconds) was included before the response screen appeared, in order to prevent the answer options from masking the stimuli. Once participants pressed the key that corresponded to the location, the next trial began, starting with the placeholders screen. Importantly, throughout the experiment and in contrast with Chen and Wyble (2015), there were only two possible target letters (‘A’ or ‘E’). The location of the target was equally distributed among the four possible locations.

This localization-task continued for 32 trials, after which they unexpectedly received two forced-choice questions on the 33rd trial, immediately after the inter-stimulus-interval. The surprise questions required them to first indicate the identity and then the color of the letter they had just seen, by pressing the corresponding key (‘5’, ‘6’, ‘7’, or ‘8’). Contrasting Chen and Wyble (2015), the identity surprise question included two novel letters as answer options (‘M’ and ‘T’), besides the previously presented letters (‘A’ and ‘E’). The color surprise question, on the other hand, contained four colored lines that represented the four colors that had been shown throughout the experiment. Considering the fact that the color is irrelevant for both the generation of a response (indicating the location of the letter) and for separating the target from the distractors, the color-task served as a control. For the identity-task as well as the color-task, the response options were horizontally positioned on the screen with the corresponding keys below them. The location of the letters and the colored lines were randomized, with each having an equal probability of appearing at one of the four locations. Additionally, the presentation of the target letter on the surprise trial was counterbalanced across participants.

Finally, after the color-task, participants were asked to indicate the location of the target letter. The structure of the surprise trial was then repeated over the next four trials, making the total number of trials 37.

Figure 3

A Schematic Overview of The Current Experiment


Placeholders (800 - 1800 ms)
Stimulus Screen (250 ms)
A
3
5
4
Blank Screen (400 ms)
Response Screen
Press a corresponding number to indicate the "Location" of the target letter
1
3
2
4
Surprise! What was the color of the target? Press the corresponding number.
5
6
7
8
Surprise! What was the target identity? Press the corresponding number.
5
6
7
8
T
E
M
A
Color Task
Identity Task

Note. The images are not true to size, but have been enhanced to improve visibility.

Data Analysis

All analyses have been performed in R (Version 4.1.1; Core Team, 2021) and the ggplot2 package (Wickham, 2016) was used for data visualization. For one participant, the entire experiment was repeated seven times due to a bug in the code. As such, the data obtained after this participant completed the first run of the experiment (i.e. the first 37 trials), was removed from the analyses.

Considering the fact that participants contribute to more than one cell when comparing performance on the surprise trial to the subsequent control trials, we performed McNemar's test (McNemar, 1947) to ascertain whether the difference in performance was significant. Similarly, to ascertain whether performance on the surprise trial was significantly better than chance level (25%), we performed a binomial test for each task.

Moreover, we expected the participants to be poor at reporting the exact identity on the identity surprise question. However, we also expect their incorrect answers to be biased toward the previously presented letters. To analyze the latter, we performed two multinomial analyses on the incorrect responses for the identity-task on the surprise trial using the EMT package (Version 1.2; Menzel, 2021). The first assumed a uniform distribution based on chance-performance (i.e. 25% for each answer option), and tested this hypothesized model on the obtained results. The second assumed a non-uniform distribution that was biased toward the previously presented letters by including pseudo-counts, and similarly assessed the goodness-of-fit for the obtained results. The pseudo-counts reflected a theoretical internal distribution participants might construct if they would count each letter they had seen. However, this would imply that the novel letters ('M' & 'T') would receive zero counts, which is mathematically complicated when calculating probabilities. As such, all four letters were first dealt one count. The remaining number of incorrect answers for the identity-task on the surprise trial were then equally distributed over the two letters that had been shown throughout the experiment.

We expected the distribution of errors for the identity-task to be centered around the previously presented letters and therefore hypothesize that this distribution is not uniformly distributed.

Data and Code Availability

All data, including the scripts used for data analysis and programming the experiment, are available online at OSF (https://osf.io/8myv9/).

Results

Attribute Amnesia

On the pre-surprise trials, the average accuracy for correctly identifying the location of the target letter was 88%. This indicates that the target letter was clearly visible, and that participants have seen the target letter. However, as we expected, only 9 out of 22 participants (41%) were able to correctly report the identity of the target letter on the surprise trial, suggesting that we indeed replicated the attribute amnesia effect. This performance was not significantly better than chance level performance (41% vs. 25%, p = 0.075). Nevertheless, on the control trial immediately after the surprise trial, the accuracy for correctly identifying the target letter's identity rose dramatically (77%). To test whether the performance on the surprise trial and the first control trial was significantly different, we employed McNemar's test (McNemar, 1947). The results from the test were significant (χ^2(1, N = 22) = 5.333, p = 0.021), suggesting that the performance for the identity-task was significantly better on the first control trial compared to the surprise trial. This rise in accuracy continued throughout the three remaining control trials (See Figure 4A).

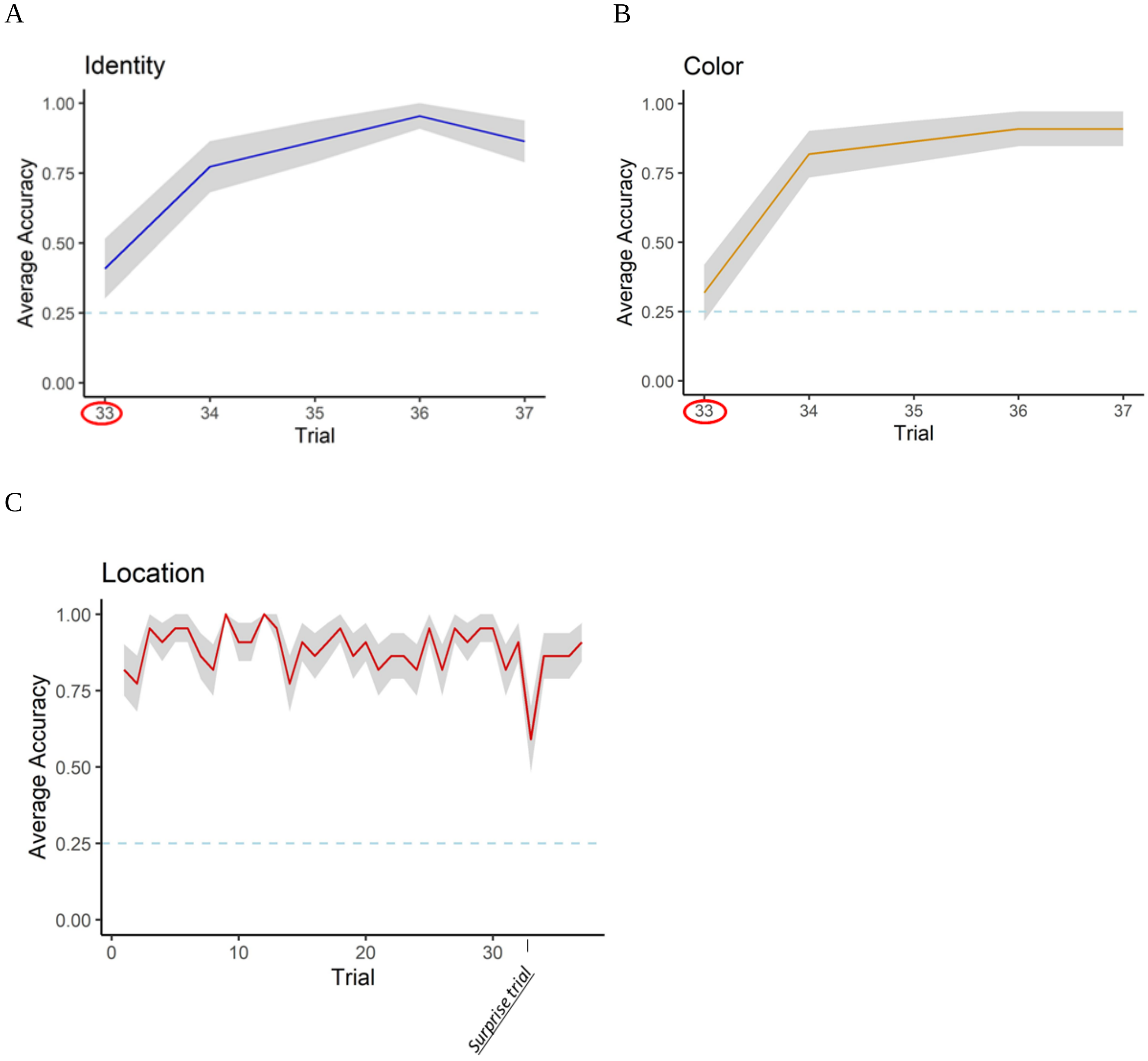
Similarly and as expected, only 7 out of 22 participants (32%) correctly reported the color of the target letter on the surprise trial, which once more indicates we replicated the attribute amnesia effect. Analogous to the performance on the identity-task, the performance on the color-task was not significantly better than chance level (32% vs. 25%, p = 0.301). However, as can be seen in Figure 4B, the color-task accuracy surged on the first control trial and differed significantly from the color-task performance on the surprise trial (82% vs. 32%, (χ^2(1, N = 22) = 11.000, p < .001).

Although there was a slight drop in performance for the location-task on the surprise trial, it quickly recovered on the following control trials (Figure 4C), which suggests that participants were able to perform the task when they expected to do so. Importantly, with 13 out of 22 participants correctly reporting the location on the surprise trial, the performance was above chance level (59% vs. 25%, p < .001).

Overall, these results confirm our hypothesis that participants are not only poor at correctly reporting a task-irrelevant attribute (color), but also exhibit a striking inability to report the target-defining attribute (identity) when there is no expectation to do so.

Figure 4

Results from The Identity-Task (A), The Color-Task (B) and The Location-Task (C).



Note. The average accuracy in subfigures A and B refer to the proportion of participants that correctly answered the surprise question, whereas in subfigure C the average accuracy refers to the average proportion of correctly identified locations for all participants. The gray area denotes the standard error of the mean in all three subfigures. Additionally, the dashed blue line represents chance-performance.

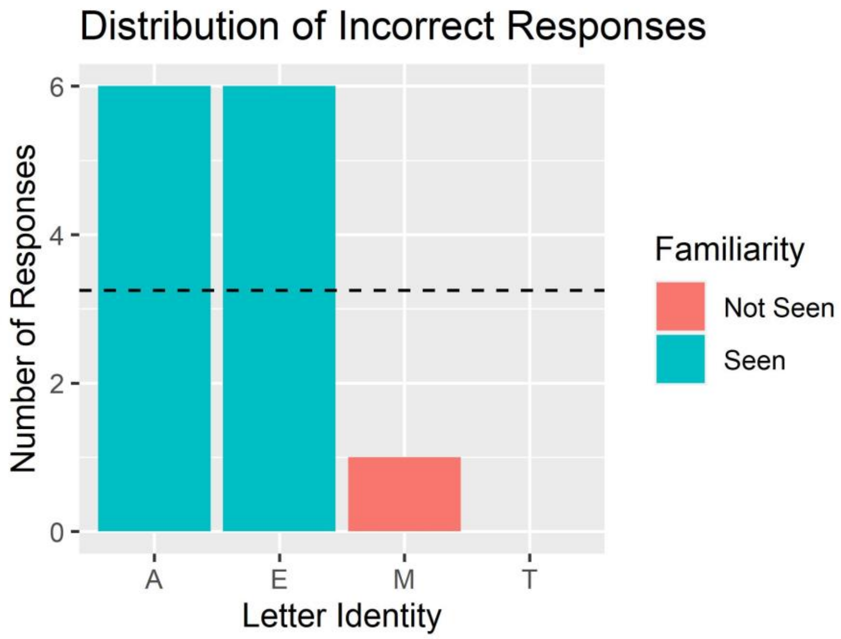
Response Bias

Taking a closer look at the incorrect responses for the identity-task on the surprise trial revealed that nearly all participants (92%) chose one of the previously shown target letters (‘A’ or ‘E’), rather than one of the two novel letters (‘M’ and ‘T’). In fact, from the 13 participants who incorrectly reported the identity of the letter on the surprise trial, only one participant chose one of the novel letters (Figure 5). The proportion of incorrect responses therefore appears to follow a non-uniform distribution. This was confirmed by the results from the multinomial test, which indicated that the hypothesized uniform distribution should be rejected ($\chi^2(1, N = 22) = 9.462, p = 0.029$). Likewise, the results from the multinomial analysis that assumed a non-uniform distribution indicated that this distribution cannot be rejected ($\chi^2(1, N = 22) = 1.091, p = 0.886$), which provided further evidence that the incorrect responses for the identity-task on the surprise trial follow a non-uniform distribution.

Together with the observed attribute amnesia for the target letter’s identity and color, these findings confirm our hypothesis that participants know they have not seen the two novel letters (‘M’ and ‘T’), yet, due to attribute amnesia, are unable to report the exact letter they have seen.

Figure 5

The Distribution of Errors for The Identity-Task



Note. The dashed black line visualizes the uniform distribution based on chance-performance.

Discussion

Prior research has indicated that, despite the occurrence of attribute amnesia, some residual information is explicitly reportable and might be automatically gathered via statistical learning (O'Donnell et al., 2021). However, with regard to stimulus identity the literature provides conflicting evidence, with some suggesting attribute amnesia is driven by the fact that stimulus attributes are not encoded at all if the current task does not require them to (Chen & Wyble, 2016), and others who suggest that such attributes are encoded, but not bound to a specific trial (Chen & Howe, 2017). To gain more clarity on what is forgotten in attribute amnesia, we used an adapted version of Chen and Wyble's (2015) paradigm to investigate whether stimulus identity is automatically encoded into an explicitly retrievable statistical representation of the previously attended stimulus identities. Taking a different approach than Chen and Wyble (2015), we only showed two target letters throughout the experiment and included two novel letters as answer options on the surprise trial.

First of all, we found that participants were poor at reporting both the target letter's identity (41%) and color (32%) when they did not expect to do so. These results are similar to the findings of Chen and Wyble (2015) and suggest that having seen a stimulus is not sufficient to make its memory trace available for immediate report. More importantly, our results provide strong evidence that participants gradually build a statistical representation of the attended stimulus identities. Though participants are poor at identifying the exact identity they just saw, their responses indicate that they do remember an average of the overall attended identities, as most participants mistakenly chose either one of the previously presented letters. Furthermore, the average accuracy of 41% on the identity-task is in line with the expected internal distribution of probabilities for each letter, which would theoretically be 0.5 for each of the two attended identities. An additional binomial test revealed that the obtained accuracy and the hypothetical distribution do not differ significantly (41% vs. 50%, $p = 0.524$). However, the obtained accuracy was not found to be significantly better than chance level performance. Future research could further explore the probabilistic nature of the internal model by using a larger sample size and manipulating both the number of targets and their frequency, as both appear to influence statistical learning (Brady et al., 2009).

Taken together, our results suggest that participants do indeed automatically extract statistical information about the stimulus identity, and are able to use this internal model to their advantage. Similar to O'Donnell et al. (2021), participants were able to make a distinction between what they had and had not seen. An alternative explanation for this can be found in the expectation-based binding hypothesis (Chen et al., 2016). According to this theory, participants would have long-term memory traces for the familiar letters ('A' and 'E'), but not for the novel letters ('M' and 'T'). When probing both familiar and novel stimuli simultaneously, the existence of such a memory trace would allow participants to make a distinction between what they have seen before and what not. Nevertheless, considering the fact that the memory traces are formed irrespective of their spatial presentation (i.e. trial) when the information is not expected to be useful, attribute amnesia would still occur when a distinction has to be made between familiar stimuli (Chen et al., 2016; Chen & Howe, 2017). Our results appear to support that notion. Additionally, the expectation-based binding hypothesis (Chen et al., 2016) can also explain the increased accuracy in the control trials after the surprise trial. Namely, because the stimulus attributes are now expected to be relevant, their memory traces are bound to their respective trial, which would facilitate performance on the identity-task and color-task.

However, the fact that Chen et al. (2019a, Experiment 2a) still obtained an attribute amnesia effect despite three of the four answer options being completely novel on the surprise trial, highlights the influence of repetitiveness of the targets on the ability to make a distinction between previously attended and novel stimuli. Based on the results from the current experiment, the results from Chen et al. (2019a, Experiment 2a) could be explained by the lack of regularity in the presented information, which could have prevented participants from constructing a reliable internal model via statistical learning. To shed more light on the possibility that participants do indeed construct an internal statistical representation of the attended stimulus identities, future research could manipulate the repetitiveness of the target. For instance, similar to the current experiment, two target letters could be shown for a duration of 31 trials. Then on the trial prior to the surprise and on the surprise trial itself, two novel letters could be introduced as targets. Subsequently, the answer options for the surprise trial would contain not only the two unique letters, but also the two letters that have been shown repeatedly. If participants indeed create an internal model via statistical learning, then their responses on the identity-task would be biased toward either one of the two repeated letters (de Lange et al., 2018).

Lastly, our results provide important insights into how objects are represented in memory. The fact that the response-irrelevant stimulus attribute (identity) appears to be remembered irrespective of task demands, supports the object-based encoding theory (Marshall & Bays, 2013). According to this theory, each object feature is automatically stored, irrespective of its relevance. Our results then suggest that the nature of this representation might reflect participants' internal model of the attended objects. Additionally, prior research has shown that even stimulus attributes that are both response-irrelevant and non-defining, such as color in the current experiment, appear to leave an implicit memory trace (Harrison et al., 2021; Shin & Ma, 2016; Born et al., 2020; Swan et al., 2016). To further explore both the object-based encoding theory and existence of an internal statistical model, future research could replicate the current paradigm, and include two novel colors when probing the target letter's color. Because this attribute is completely irrelevant in the current experiment, the ability to distinguish between familiar and novel colors would show whether object-based encoding relies on an internal statistical model of the stimulus' features.

To conclude, the current study showed that not all information regarding the stimulus identity is lost in attribute amnesia. In fact, it appears that stimulus identity is automatically encoded by means of statistical learning, which subsequently leads to the gradual creation of an internal model containing an average of the attended identities. Importantly, despite the fact that participants were unable to report the exact identity, they were able to use this internal model to their advantage when a distinction needed to be made between familiar and novel stimuli, which implies that this information is encoded in an explicitly retrievable form. Therefore, attribute amnesia might reflect the inability to make a distinction between information within the gradually constructed internal model of the attended stimuli.

References

Born, S., Jordan, D., & Kerzel, D. (2020). Attribute amnesia can be modulated by foveal presentation and the pre-allocation of endogenous spatial attention. *Attention, Perception, & Psychophysics, 82*(5), 2302–2314. https://doi.org/10.3758/s13414-020-01983-7

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology: General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

Chen, W., & Howe, P. D. L. (2017). Attribute amnesia is greatly reduced with novel stimuli. *PeerJ*, 5, e4016. https://doi.org/10.7717/peerj.4016

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. https://doi-org.proxy-ub.rug.nl/10.1016/j.cognition.2015.11.014

Chen, H., & Wyble, B. (2015). Amnesia for object attributes: failure to report attended information that had just reached conscious awareness. *Psychological Science*, *26*(2), 203–10. https://doi.org/10.1177/0956797614560648

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*(2), 225–34. https://doi.org/10.1037/xhp0000133

Chen, H., Yu, J., Fu, Y., Zhu, P., Li, W., Zhou, J., & Shen, M. (2019a). Does attribute amnesia occur with the presentation of complex, meaningful stimuli? The answer is, “it depends”. *Memory & Cognition*, 47(6), 1133–1144. https://doi.org/10.3758/s13421-019-00923-7

Chen, H., Yan, N., Zhu, P., Wyble, B., Eitam, B., & Shen, M. (2019b). Expecting the unexpected: violation of expectation shifts strategies toward information exploration. *Journal of Experimental Psychology. Human Perception and Performance*, *45*(4), 513–522. https://doi.org/10.1037/xhp0000622

Harrison, G. W., Kang, M., & Wilson, D. E. (2021). Remembering more than you can say: re-examining “amnesia” of attended attributes. *Acta Psychologica*, *214*. https://doi.org/10.1016/j.actpsy.2021.103265

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), 1331–7. https://doi.org/10.1037/xlm0000231 de Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, *13*(2). https://doi.org/10.1167/13.2.21

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences [Computer software]. *Behavior Research Methods*, *44*(2), 314-324. doi:10.3758/s13428-011-0168-7

McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, *12*(2), 153–157. https://doi.org/10.1007/bf02295996

Menzel, U. (2021). EMT Package [Computer software]. Retrieved from https://CRAN.R-project.org/package=EMT

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 411–421. http://dx.doi.org/10.1037/0278-7393.28.3.411

O'Donnell, R. E., Chen, H., & Wyble, B. (2021). No explicit memory for individual trial display configurations in a visual search task. *Memory & Cognition*, *49*(8), 1705–1721. https://doi.org/10.3758/s13421-021-01185-y

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

Sherman, B. E., Graves, K. N., & Turk-Browne, N. B. (2020). The prevalence and importance of statistical learning in human cognition and behavior. *Current Opinion in Behavioral Sciences*, *32*, 15–20. https://doi.org/10.1016/j.cobeha.2020.01.015

Shin, H., & Ma, W. J. (2016). Crowdsourced single-trial probes of visual working memory for irrelevant features. *Journal of Vision*, *16*(5), 10–10. https://doi.org/10.1167/16.5.10

Swan, G., Collins, J., & Wyble, B. (2016). Memory for a single object has differently variable precisions for relevant and irrelevant features. *Journal of Vision*, *16*(3), 32–32. https://doi.org/10.1167/16.3.32

Swan, G., Wyble, B., & Chen, H. (2017). Working memory representations persist in the face of unexpected task alterations. *Attention, Perception & Psychophysics*, *79*(5), 1408–1414. https://doi.org/10.3758/s13414-017-1318-5

Tam, J., Mugno, M. K., O'Donnell, R. E., & Wyble, B. (2021). And like that, they were gone: a failure to remember recently attended unique faces. *Psychonomic Bulletin & Review*, *28*(6), 2027–2034. https://doi.org/10.3758/s13423-021-01965-2

Wang, R., Fu, Y., Chen, L., Chen, Y., Zhou, J., & Chen, H. (2021). Consciousness can overflow report: novel evidence from attribute amnesia of a single stimulus. *Consciousness and Cognition*, *87*, 103052–103052. https://doi.org/10.1016/j.concog.2020.103052

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis [Computer Software]. Retrieved from https://ggplot2.tidyverse.org