PNAS

Supporting Information for
Repetition Learning Is neither a Continuous nor an Implicit Process

*Philipp Musfeld, Alessandra S. Souza, Klaus Oberauer

*Philipp Musfeld
Email: philipp.musfeld@psychologie.uzh.ch

This PDF file includes:

Supporting text
Figures S1 to S3
Tables S1 and S2
SI References

Supporting Information Text

SI Methods

Participant Exclusion Criteria

Data for this study was collected online via the online participant platform Prolific. To ensure high quality of the collected data, multiple in- and exclusion criteria were defined.
On Prolific, we applied the following criteria for participation: age between 18 and 35 years old, fluent in English language, normal or corrected to normal vision and no color blindness. We also restricted participation to participants who did not participate in any other Hebb experiments we had collected on Prolific.
After reading the instructions and before starting with the main part of the experiment, participants had to answer four questions about the experiment to make sure that all relevant information was understood. Participants were only allowed to participate if all questions were answered correctly. When failing the instruction check for the first time, participants were redirected to the instruction pages and given a second chance. In case of two consecutive failures, the study was ended.
During the study, we controlled that participants stayed on the experiment screen by tracking the visibility of the browser window. If the current browser window was hidden more than three times, participants were excluded from the study.
After completion of the study, participants were asked if they 1) participated seriously in the study and 2) used any aids to help improve their performance during the study (cheating). Participants were only considered for the data analysis if they indicated a serious participation and no use of aids. Additionally, we excluded participants from the data analysis whose performance in the working memory task was at chance level during at least one half of the experiment. Chance level performance was assessed by comparing participants' average performance in Filler trials in one half of the experiment to the 99% quantile of the binomial distribution, with guessing probability set to one divided by the number of possible responses (verbal: 1/18, visual: 1/9) and number of events set to the number of responses within one half of the experiment (verbal: 270; visual: 360). For the verbal experiment, this resulted in a threshold of at least 8.5% correct responses, and for the visual experiment to a threshold of at least 16% correct responses.
Overall, these criteria led to the exclusion of 3 participants in the verbal experiment and 27 participants in the visual experiment.

Creation of Memory Sets

Verbal Experiment. Memory sets in the verbal experiment consisted of 9 consonants. Consonants were sampled from the set of all consonants except for "W" and "Y" because "W" is the only consonant containing two syllables and "Y" has the same function as a vowel in English language. Consonants were presented in a row of nine boxes in the center of the screen. Boxes were quadratic with widths and height set to 14% of the vertical screen height. Font size of consonants was set to 8% of the vertical screen height.
All consonant lists were generated randomly anew upon starting the experiment but with the following constraints: 1) All stimulus lists needed to differ in at least three item-position combinations to avoid random duplication of Filler lists; 2) Filler lists were not allowed to share the first two consonants with the Hebb list; 3) Lists were not allowed to contain well known acronyms to avoid effects of semantic chunking. For this, a list of 81 well know 2- and 3-letter acronyms was collected (e.g., "PDF", "BMW", "CV") and acronyms were not allowed to be part of any lists; 4) The consonants "M" and "N" were not allowed to be part of the same list because of their high degree of phonemic similarity.

Visual Experiment. Memory sets in the visual experiment consisted of 6 colored squares (square side corresponding to 50x50 viewport scaled pixels) presented at random locations against a grey (RGB 128 128 128) background. Colors were selected from a set of nine discrete colors (RGB): white (255 255 255), black (0 0 0), blue (0 0 255), cyan (0 255 255), green (0 255 0), yellow (255 255 0), orange (255 128 0), red (255 0 0), and magenta (255 0 255). Locations were selected from a set of 49 locations spanned by a 7x7 invisible grid (grid cell size corresponding to 50x50 viewport scaled pixels) centered in the middle of the screen. Again, all visual arrays were generated randomly anew upon starting the experiment but with the following constraints: 1) All squares within an array were separated by at least one grid cell so that squares were not allowed to touch each other; 2) All created arrays had to differ in at least two color-location combinations.

Bayesian Hierarchical Mixture Modeling:

All models described in the following section are Bayesian hierarchical mixture models. The hierarchical implementation of the models allows to estimate both, population-level parameters to describe the sample as a whole, and individual-level-parameters for each participant individually. The mixture structure of the models incorporated the assumption that the observed data was generated by different data generating processes. We describe this mixture process in more details below when explaining the models. The model script for the exact implementation of all models is available at https://osf.io/dpkyb/ (1).

Modeling of Working Memory Data. For the working memory task, we assumed that the observed data could have been produced by one of two different generative processes: a learning process in which memory for the repeated Hebb set improves over trials, or a non-learning process in which no learning effect for the repeated Hebb set is produced. For both components of the model, we assumed that the number of correctly recalled items on each trial followed a binomial likelihood with latent parameter θ, leading to the two following likelihood components:

$$Non - Learning \sim Binomial\left(successes_j \,\middle|\, trials_j, \theta_{non-learning_{i,j}}\right)$$
$$Learning \sim Binomial\left(successes_j \,\middle|\, trials_j, \theta_{learning_{i,j}}\right)$$

In both models, the latent parameter $\theta_{i,j}$ reflects the ability of the *ith* participant to recall the current memory set in trial *j*. The learning and the non-learning process differed in how $\theta_{i,j}$ was modeled.

For the non-learning process, we assumed that the ability of recalling a memory set does not differ between Filler sets and Hebb sets over the experiment. Therefore, we modeled $\theta_{non-learning}$ as a linear function of the mini-blocks of trials without distinguishing between Filler and Hebb sets. Each mini-block included the presentation of one repetition of the Hebb set and three unrepeated Filler sets. The linear effect of mini-block allowed the model to account for slight changes in memory performance over the experiment which might be caused by fatigue or practice effects. To make sure that $\theta_{non-learning}$ is a probability, a logit-link was applied. This led to the following model equation for $\theta_{non-learning}$

$$\theta_{non-learning_{i,j}} = logit^{-1}\left(\alpha_i + \beta_{block_i} * block_j\right)$$

with α_i reflecting the intercept of the *ith* participant and β_{block_i} reflecting the slope of the linear effect of mini-block for the *ith* participant in trial *j.*

For the learning process, we assumed that, at a certain point in the experiment, a person's ability to recall the repeated Hebb set improves over repetitions compared to the ability to recall an unrepeated Filler set. For this, we used the non-learning model as a baseline to describe participants' working memory performance in unrepeated Filler sets and before learning had started. On top of this, we added a linear term which allowed performance on the repeated Hebb set to increase after an estimated onset point of learning. Again, a logit-link was applied to make sure that $\theta_{learning}$ is a probability.

$$\theta_{learning_{i,j}} = logit^{-1}(\alpha_i + \beta_{block_i} * block_j + setType_j * \beta_{learning_i} * \max(0, block_j - \beta_{onset_i}))$$

In this equation, $setType_j$ codes if the presented memory set in trial *j* was a Filler set (= 0) or the Hebb set (= 1), and $\beta_{learning_i}$ reflects the rate of the learning effect of participant *i* on the repeated Hebb set. The "max" function allows to offset the onset of learning on the time scale of the experiment so that no learning benefit is added before the onset of learning is reached. The onset of learning for each participant *i* is reflected in β_{onset_i}.

In the formula stated above, the model predicts that performance on the Hebb set approaches perfection once the learning process has started (upper asymptote of 1). Although this is in principle plausible once the Hebb set was learned, the hard constraint on the upper asymptote can still cause sampling problems when participants make mistakes on the Hebb set after learning it (e.g., by clicking the wrong button or having an attentional lapse). To account for this sort of errors, we loosened the constraint on the upper asymptote and introduced an additional parameter which estimates the upper asymptote in the boundaries between 0.85 and 1:

$$\theta_{learning_{i,j}} = \beta_{asymtote} * logit^{-1}(\alpha_i + \beta_{block_i} * block_j + setType_j * \beta_{learning_i} * \max(0, block_j - \beta_{onset_i}))$$

Given the non-learning and the learning part of the model, we first computed the likelihood for each participant under both models:

$$L_{non-learning_i} = \prod_{j=1}^{ntrial} Binomial(successes_j \mid trials_j, \theta_{non-learning_{i,j}})$$

$$L_{learning_i} = \prod_{j=1}^{ntrial} Binomial(successes_j \mid trials_j, \theta_{learning_{i,j}})$$

Given the two likelihood components for each participant, we then computed the mixture likelihood with mixing proportion λ on the participant level for each participant *i* as

$$L_i \sim \lambda * L_{learning_i} + (1 - \lambda) * L_{non-learning_i}$$

The mixture proportion λ is applied on the participant level and therefore indicates the proportion of learning participants in the full sample. The posterior probability of belonging to the learning or the non-learning process for a single participant can be recovered by the following equation:

$$p_{learning_i} = \frac{L_{learning_i} * \lambda}{L_{learning_i} * \lambda + L_{non-learning_i} * (1 - \lambda)}$$

For fitting the model, all data variables were scaled prior to modeling so that all parameters could be fitted roughly on unit scale. The indicator for the set type was dummy coded with Filler = 0 and Hebb = 1. The mini-block variable was scaled into a range of [0, 1]. Consequently, the parameter for the onset of learning β_{onset} was also restricted to the range of [0, 1]. We further restricted the position of the onset point to a maximum of the last minus one mini-block to ensure identifiability of the two model components. Otherwise, the learning model could, in principle, mimic the non-learning model by setting the onset of the learning curve to the very last data point. A similar issue applies to the parameter for the learning rate $\beta_{learning_i}$. This parameter needs to be larger than 0 to ensure identifiability of the two model components. Here, the lower boundary of this parameter was set to 3, which is an arbitrary choice based on visual explorations of the implemented model function. Additionally, we set an upper boundary of 500 to this parameter which roughly corresponds to the point at which the learning curve approaches a step function. All bounded parameters were logit-transformed to allow the sampler to operate more efficiently in an unbounded space and to estimate parameters roughly on unit scales. Weakly informative priors following the recommendations from the Stan Development Team were used for all parameters. The exact prior specifications are shown in Table S1.

Modeling of Awareness Ratings. The same modeling approach described above was applied to the data of the awareness rating task. Again, we assumed that the data could have been produced by one of two different data generating processes: an aware process in which participants become aware of the repetition of the Hebb set, or a not-aware process in which participants do not become aware of the repetition. The only difference from the memory model is that another likelihood function is needed because awareness ratings were assessed on a visual slider scale. Defining an appropriate likelihood function for data obtained from a visual slider scale, however, is not as straightforward. This is because this sort of data can be treated as a continuous measure, but it is bounded by the limits of the scale. Therefore, continuous but unbounded distributions like the normal distribution do not provide an appropriate description of the data generating process as it can lead to predictions outside the boundaries of the scale. However, the use of bounded continuous distributions like the Beta distributions has also found to be problematic, as it is only defined for values between the boundaries but not for values which are equal to the boundaries (i.e., 0 and 1). Recently, a new approach for modeling this sort of the data has been proposed which is the ordered beta model (2). The ordered beta model combines the beta distribution (which handles continuous outcomes between the boundaries) with ordered cut points. These ordered cut points define at which point the bounds of the scale become more likely than the continuous values in between, and hence when the observed response would lie on the scale boundaries (in our case, 0 or 1). Unlike other approaches, like the zero-one-inflated beta model, this approach comes with the benefit that the latent variable of interest is still only reflected in a single parameter, which here is θ_{aware} and $\theta_{not-aware}$ respectively. In the ordered beta model, the two likelihood components of the model are defined as follows:

$$L_{not-aware_i} \sim \prod_{j=1}^{ntrial} \left\{ \begin{matrix} 1 - logit^{-1}\left(logit\left(\theta_{not-aware_{i,j}}\right) - k_1\right) & ; y_j = 0 \\ \left[logit^{-1}\left(logit\left(\theta_{not-aware_{i,j}}\right) - k_1\right) - logit^{-1}(logit\left(\theta_{not-aware_{i,j}}\right) - k_2)\right] * Beta(y_j \mid \theta_{not-aware_{i,j}}, \phi) & ; y_j \, \epsilon(0,1) \\ logit^{-1}\left(logit\left(\theta_{not-aware_{i,j}}\right) - k_2\right) & ; y_j = 1 \end{matrix} \right\}$$

$$L_{aware_i} \sim \prod_{j=1}^{ntrial} \left\{ \begin{matrix} 1 - logit^{-1}\left(logit\left(\theta_{aware_{i,j}}\right) - k_1\right) & ; y_j = 0 \\ \left[logit^{-1}\left(logit\left(\theta_{aware_{i,j}}\right) - k_1\right) - logit^{-1}(logit\left(\theta_{aware_{i,j}}\right) - k_2)\right] * Beta(y_j \mid \theta_{aware_{i,j}}, \phi) & ; y_j \, \epsilon(0,1) \\ logit^{-1}\left(logit\left(\theta_{aware_{i,j}}\right) - k_2\right) & ; y_j = 1 \end{matrix} \right\}$$

The likelihood for each model is split into three parts, depending on whether the observed rating *y* in a trial *j* was at the bounds of the scale (i.e., $y_j = 0$ or $y_j = 1$) or somewhere between the bounds (i.e., $y_j \, \epsilon(0,1)$). The latent variable $\theta_{i,j}$ reflects the strength of the familiarity signal which is elicited in participant *i* by the memory set presented in trial *j* and thereby determines the rating which is chosen on the slider. The parameters k_1 and k_2 reflect the ordered cut points at which the strength of the familiarity signal would translate into a discrete 0 or 1 response on the slider, with $k_1 < k_2$. The two cut points are estimated for the whole sample and do not vary between participants.

From here on, θ_{aware} and $\theta_{not-aware}$ were modeled in the same way as described for the learning model. For the not-aware process we assumed that participants do not distinguish between Hebb- and Filler-sets, because they don't recognize the repetition. Therefore, $\theta_{not-aware}$ is again modeled as:

$$\theta_{not-aware_{i,j}} = logit^{-1}\left(\alpha_i + \beta_{block_i} * block_j\right)$$

For the aware process, we assumed that participants would, at a certain point, recognize the repeated Hebb set, because a higher familiarity signal is elicited compared to the non-repeated Filler sets. This should lead to an increase in a participant's rating scores after they became aware. For that case, θ_{aware} was again modeled with the same equation as described in the learning model:

$$\theta_{aware_{i,j}} = \beta_{asymtote} * logit^{-1}\left(\alpha_i + \beta_{block_i} * block_j + setType_j * \beta_{learning_i} * \max\left(0, block_j - \beta_{onset_i}\right)\right)$$

Again, we first calculated the likelihood under both model components for each participant according to the likelihood function of the ordered beta model as specified above and then computed the mixture likelihood with mixing proportion λ with

$$L_i \sim \lambda * L_{aware_i} + (1 - \lambda) * L_{not-aware_i}$$

Here, the mixture proportion λ defines the proportion of participants in the sample who became aware of the repetition. By using the same equation as shown above, we can recover the posterior probability of belonging to the aware or not-aware process for each participant with

$$p_{aware_i} = \frac{L_{aware_i} * \lambda}{L_{aware_i} * \lambda + L_{not-aware_i} * (1 - \lambda)}$$

Again, all data variables were scaled in the same way as described for the learning model. Additionally, awareness ratings were scaled to a range of [0,1] to fit them with the ordered beta model.

Combined Analysis of Learning and Awareness Data. To assess the relationship between the learning and the awareness process, both data sets were jointly modeled in a multivariate model. For this, both models described above were combined in the same model script and a joint covariance structure between the parameters from the learning and the awareness model was specified. This allowed us to estimate the correlations between the learning and the awareness process. Parameters between the two processes were directly comparable as both models included the exact same set of parameters on the same scale. The correlation results for all parameters are presented in Figure S1.

Comparison Model: Continuous Learning Process. The modeling approach described above is at odds with the common assumption of repetition learning as a continuous process which starts from the first repetition. This divergence is reflected in the β_{onset_i} parameter, which allows learning to start at any point during the experiment. To assess the evidence in favor of including this parameter in our model, we specified an alternative model in which we fixed the onset of the learning process to the first occurrence of the repeated memory set. This was achieved by removing the β_{onset_i} parameter from the formula for modeling $\theta_{learning}$:

$$\theta_{learning_{i,j}} = \beta_{asymtote} * logit^{-1}\left(\alpha_i + \beta_{block_i} * block_j + setType_j * \beta_{learning_i} * block_j\right)$$

Additionally, the constraint on $\beta_{learning_i}$ was loosened to a minimum of 0.5 so that even smaller and more gradual effects could be captured by the model while still ensuring identifiability between the learning and the non-learning model. The rest of the model was identical to the learning model described above. To perform model comparisons based on the pointwise likelihood, the expected pointwise likelihood from the two likelihood components of the mixture model for each data point was computed as follows:

$$L_{i,j} = p_{learning_i} * L_{learning_{i,j}} + \left(1 - p_{learning_i}\right) * L_{non-learning_{i,j}}$$

Here, the pointwise likelihood under the learning model ($L_{learning_{i,j}}$) and under the non-learning model ($L_{learning_{i,j}}$) gets weighted by the posterior probability of participant *i* belonging to one or the other model ($p_{learning_i}$). Finally, model comparisons were conducted by using leave-one-out (LOO) cross-validation of the expected log pointwise predictive density (ELPD) of the two models (3). Model comparisons results are displayed in Table S2.

Model Fitting and Convergence Diagnostics. All models were implemented using the statistical programming language Stan (4) and fitted with *R v4.2.1* (5) and the R-package *rstan v2.26.13* (6). Models were fitted on four chains with 1000 warmup and 2500 post-warmup iterations per chain (10000 post-warmup iterations in total). For the multivariate learning-awareness model, 4000 post-warmup iterations were run (16000 post-warmup iterations). Convergence of models was evaluated by using Stan's convergence and efficiency diagnostics for Markov chains (7). These include improved R-hat metrics as a measure of within- and between-chain variability for model parameter estimates, bulk effective sample size as a measure of sampling efficiency from the bulk of the posterior distributions and tail effective sample size as a measure of sampling efficiency from the tails of the posterior distributions. For all fitted models, we aimed for Rhats < 1.05 and bulk and tail effective sample sizes > 100 times the number of chains (= 400). Criteria were generally met for all fitted models. Only for the learning model fitted to the *No Information* group of the visual experiment and for the multivariate model fitted to the *Awareness Rating* group of the verbal experiment warnings for low bulk and tail effective samples sizes were obtained. Investigations of these issues revealed that in both cases, warnings were related to individual parameter estimates of a single participant. In the multivariate model, warnings were only related to the obtained awareness parameters for one participant. Refitting both models without the problematic participant confirmed that issues were only related to parameters of the removed participant as warnings did not occur any longer. Yet the inclusion of these participants did not affect the model estimates, and hence both participants were still considered for further analyses.

SI Results

Learning without Awareness

The results of classifying participants from the *Awareness Rating* condition into aware / not-aware and learning / non-learning showed that three participants were classified as learning without awareness. Data of these participants would therefore be diagnostic for showing an implicit learning process. The working memory and awareness rating data of these participants are plotted in Figure S2, separated into participants belonging to the verbal and the visual experiment.
For the two participants from the verbal experiment, we can see that both participants show a clear learning effect. However, by visual inspection, we can still see indications of awareness for the repeated Hebb set, which also align with the trial-by-trial performance in the working memory task. Although this was not captured in our model's classification process, a close relation between trial-by-trial memory performance and trial-by-trial awareness ratings was still observed. In contrast, the awareness ratings of the participant from the visual experiment (Example 3) clearly show no signs of awareness for the repeated Hebb set. However, here, the learning process is associated with a lot of uncertainty. Although performance for the repeated Hebb set seems to increase at the end of the experiment, there is still high variability in performance, and the range of model predictions with parameters sampled from their 95% highest density interval still includes performance levels equal to Filler trials.
Overall, none of these participants provides convincing evidence for the presence of an implicit learning effect.

Awareness without Learning

Compared to the three participants who were classified as learning without awareness, a larger number of participants was classified as aware without showing a learning effect. This was especially pronounced in the visual experiment with 22 participants falling in this category, but not so much for the verbal experiment with only 6 participants in this category. Figure S3A shows examples of the working memory and awareness rating data from four of these participants. All presented examples clearly
show that participants became aware of the repetition but did not produce any learning effect. At this point, we can only speculate about reasons why these participants did not show a learning effect although being aware of the repetition.

One possibility is that these participants just became aware of the repetition very late in the experiment so that there was not enough time left to build a stable representation of the repeated memory set. Figure S3B displays the average onset point of awareness for participants who learned the repeated memory set and those who did not. For the visual experiment, we can see that participants who did not show a learning effect indeed also showed a later onset of awareness. However, this was not the case for all non-learning participants. Additionally, no difference in the onset of awareness was found for the verbal experiment. Therefore, this explanation can only partially account for the participants who don’t learn although being aware of the repetition.

Another possibility is that these participants did not engage in learning the repeated memory set after becoming aware of the repetition. This would be consistent with our claim that the formation of a new stable knowledge structure requires an explicit strengthening process in which participants deliberately need to engage. Therefore, awareness of the repetition would be a necessary but not a sufficient condition for producing the learning effect.

SI Figures and Tables

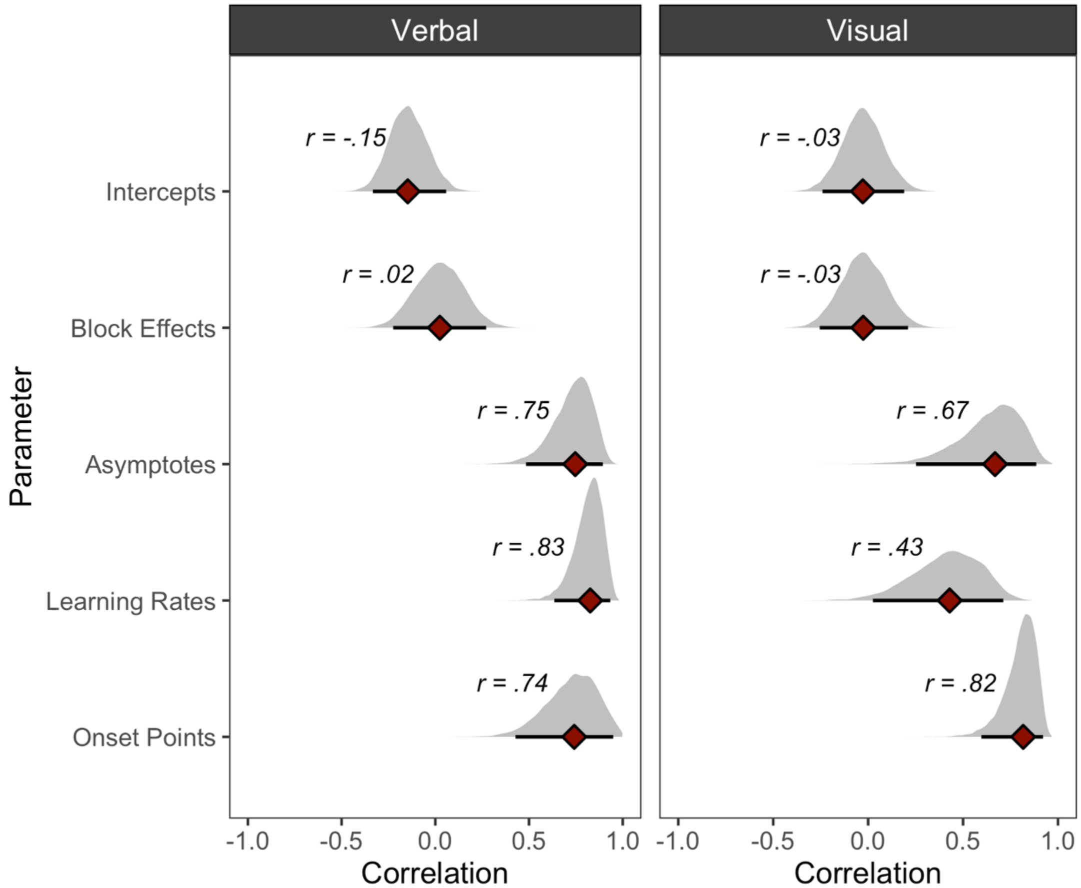

Fig. S1. Posterior distributions of the estimated correlations between the parameters of the learning model and parameters of the awareness model. Points reflect the median of the distribution, lines the 95% highest density interval.

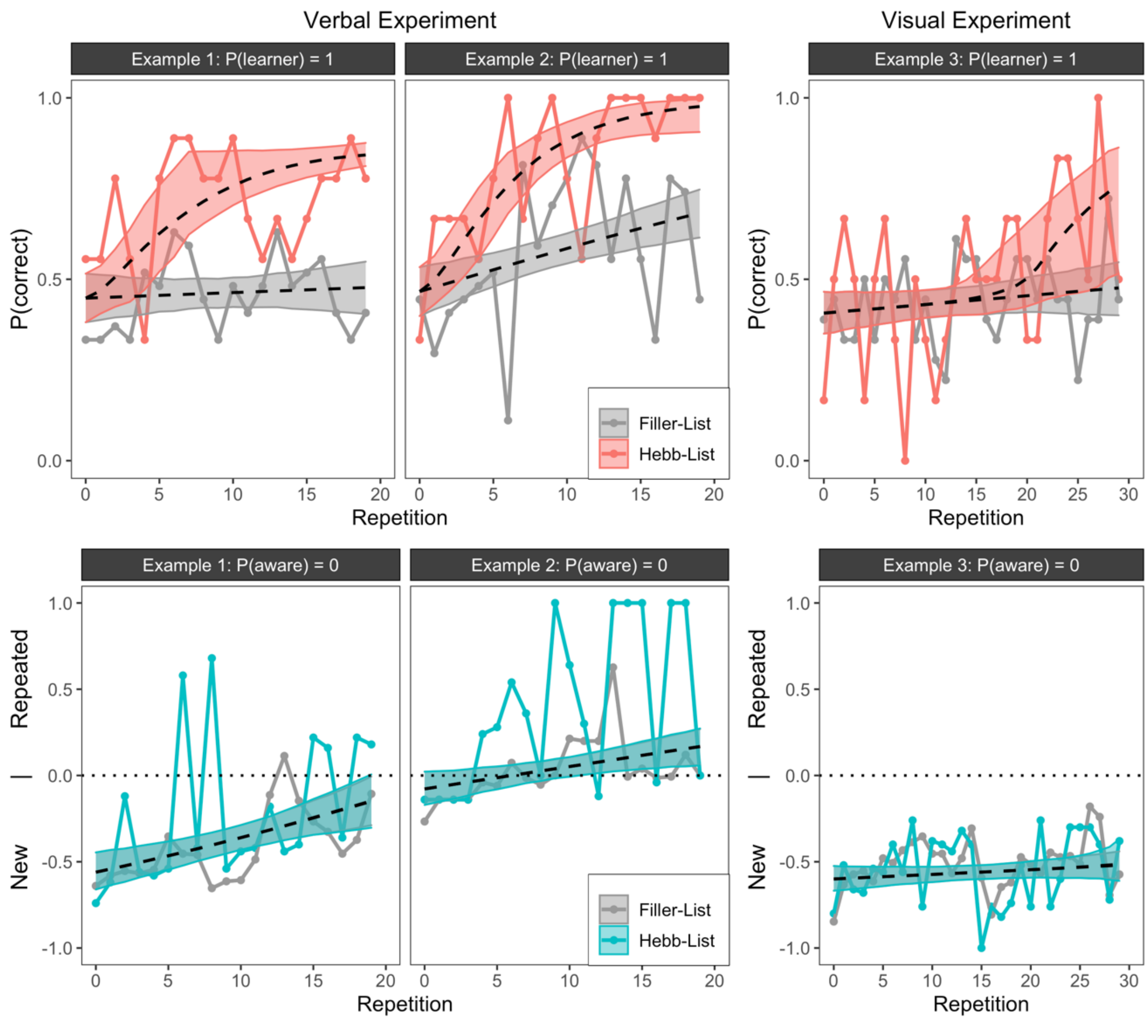

Fig. S2. Learning and awareness data of the three participants which were classified as showing a learning effect without showing awareness of the repetition. The dashed line indicates the predictions of the model with the best fitting parameters, and the colored areas indicate the range of model predictions with parameters sampled from their 95% highest density interval. *Note*: P(correct) = proportion of correct responses. The x-axes show the repetition number for the Hebb-set. For the filler sets, this corresponds to average performance in each mini-block.

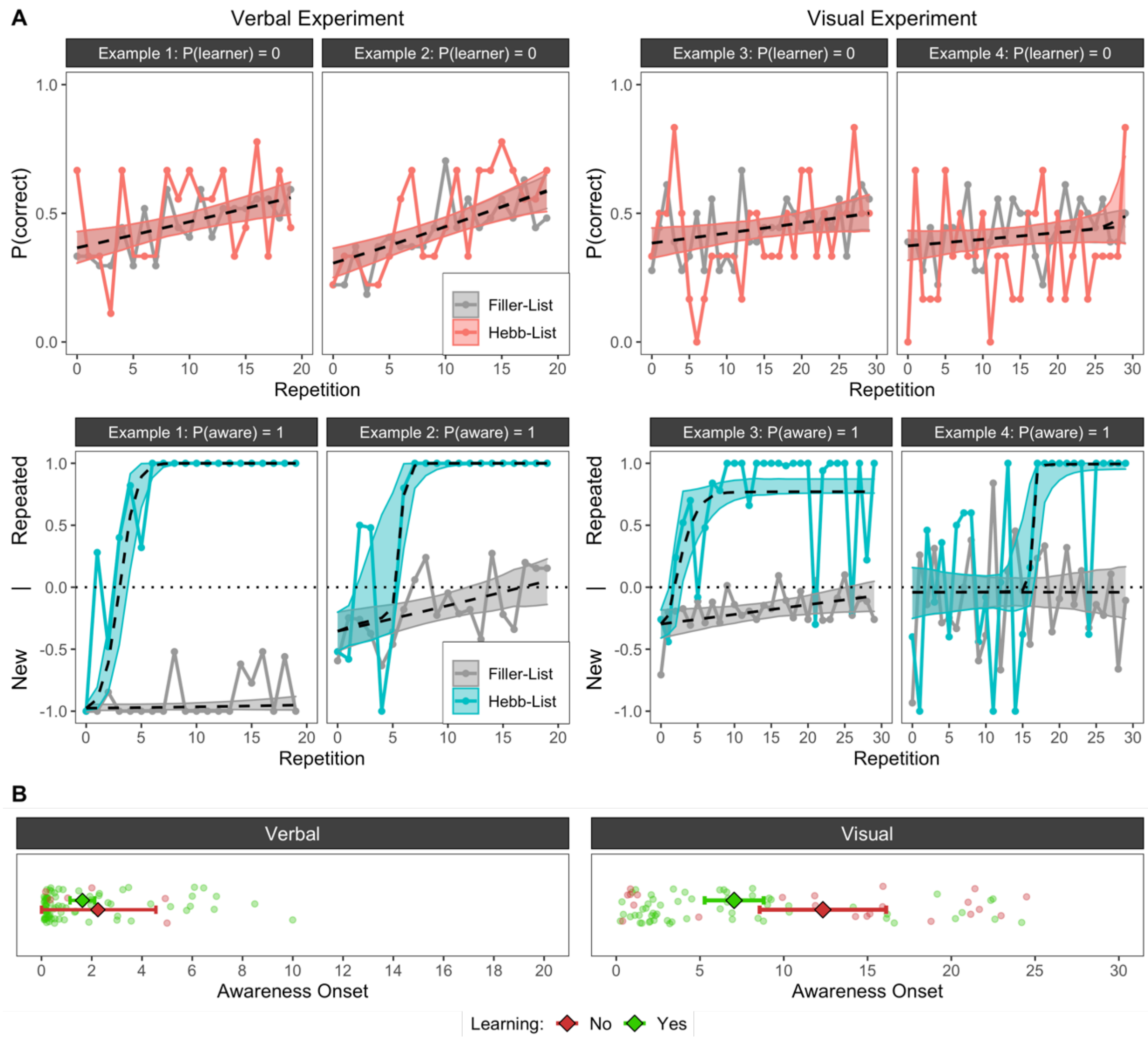

Fig. S3. A Examples of participants from both experiments showing awareness of the repetition without producing a learning effect. The dashed line indicates the predictions of the model with the best fitting parameters, and the colored areas indicate the range of model predictions with parameters sampled from their 95% highest density interval. **B** Average onset point of awareness for participants classified as learning and participants classified as non-learning. Error bars represent 95% confidence interval. Data points represent the individual onset points of awareness for all participants. *Note*. P(correct) = proportion of correct responses. The x-axes in panel A show the repetition number for the Hebb-set. For the filler sets, this corresponds to average performance in each mini-block.

Table S1. Prior specifications for all parameters used in the described models. β parameters describe the population-level effect of a parameter, σ parameters describe the standard deviation of the corresponding effect, and δ parameters describe the difference of each individual from the population-level effect. Note that for the hierarchical implementation, a non-centered parametrization was used in which an individual participant effect δ is defined by multiplying the standardized effect zδ by the standard deviation σ of the corresponding parameter. Therefore, priors are applied to zδ. If transformations were applied, priors are set on the transformed scale.

Parameter	Prior	Description	Transformed
λ	$Beta(2, 2)$	Mixture proportion	No
ϕ	$Normal(3, 1)$	Scale parameter of Beta likelihood	Log
k_1, k_2	$induced\text{-}Dirichlet(1)$	Ordered cut points for ordered beta model	No
α	$Normal(0, 1)$	Population-level effect for intercept	Logit
σ_α	$student(4, 0, 1)_{I(0,\infty)}$	Standard deviation of α	No
$z\delta_\alpha$	$Normal(0, 1)$	Standardized effect on α	No
β_{block}	$Normal(0, 1)$	Population-level effect of mini-block	Logit
$\sigma_{\beta_{block}}$	$student(4, 0, 1)_{I(0,\infty)}$	Standard deviation of β_{block}	No
$z\delta_{\beta_{block}}$	$Normal(0, 1)$	Standardized participant effect on β_{block}	No
$\beta_{learning}$	$Normal(-3, 1)$	Population-level effect for learning rate	Logit
$\sigma_{\beta_{learning}}$	$student(4, 0, 1)_{I(0,\infty)}$	Standard deviation of $\beta_{learning}$	No
$z\delta_{\beta_{learning}}$	$Normal(0, 1)$	Standardized effect on $\beta_{learning}$	No
β_{onset}	$Normal(0, 1)$	Population-level effect for onset point	Logit
$\sigma_{\beta_{onset}}$	$student(4, 0, 1)_{I(0,\infty)}$	Standard deviation of β_{onset}	No
$z\delta_{\beta_{onset}}$	$Normal(0, 1)$	Standardized effect on β_{block}	No
$\beta_{asymptote}$	$Normal(0, 1)$	Population-level effect for asymptote	Logit
$\sigma_{\beta_{asymptote}}$	$student(4, 0, 1)_{I(0,\infty)}$	Standard deviation of $\beta_{asymptote}$	No
$z\delta_{\beta_{asymptote}}$	$Normal(0, 1)$	Standardized effect on $\beta_{asymptote}$	No

Table S2. Results of the model comparisons using leave-one-out (LOO) cross-validation of the expected log pointwise predictive density (ELPD) of the two models. Negative ELPD-LOO values differences indicate a better fit of the variable-onset model compared to the continuous-learning model.

	ELPD-LOO variable-onset	ELPD-LOO continuous-learning	ELPD-LOO Difference	SE Difference
Verbal				
No Awareness	-15817.2	-15852.6	- 35.4	16.3
Awareness Only	-14697.3	-14775.7	- 78.4	19.6
Awareness Rating	-14425.5	-14443.9	- 18.4	15.6
Visual				
No Awareness	-19073.6	-19133.6	- 60.0	21.5
Awareness Only	-19275.4	-19378.6	-121.2	24.5
Awareness Rating	-18827.9	-19023.3	-195.4	25.6

SI References

1. P. Musfeld, A. Souza, K. Oberauer, Data, Materials, and Code for: Repetition Learning: Neither a Continuous nor an Implicit Process (2022) https:/doi.org/10.17605/OSF.IO/DPKYB.

2. R. Kubinec, Ordered Beta Regression: A Parsimonious, Well-Fitting Model for Continuous Data with Lower and Upper Bounds. *Polit. Anal.*, 1–18 (2022).

3. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput* 27, 1413–1432 (2017).

4. Stan Development Team, Stan Modeling Language Users Guide and Reference Manual, Version 2.30 (2022).

5. R Core Team, "R: A language and environment for statistical computing" (2022).

6. Stan Development Team, RStan: the R interface to Stan (2022).

7. A. Vehtari, A. Gelman, D. Simpson, B. Carpenter, P.-C. Bürkner, Rank-Normalization, Folding, and Localization: An Improved Rˆ for Assessing Convergence of MCMC (with Discussion). *Bayesian Anal.* 16 (2021).