# Structure transfer and consolidation in visual implicit learning

Dominik Garber* & József Fiser

Department of Cognitive Science
Center for Cognitive Computation
Central European University, Vienna, Austria

**Corresponding Author:** Dominik Garber; Central European University Private University, Quellenstraße 51, 1100 Vienna, Austria; garber_dominik@phd.ceu.edu

**Keywords:** Transfer Learning, Unsupervised Learning, Consolidation, Implicit Learning, Statistical Learning

**Abstract**

Transfer learning, the re-application of previously learned higher-level regularities to novel input, is a key challenge in cognition. While previous empirical studies investigated human transfer learning in supervised or reinforcement learning for explicit knowledge, it is unknown whether such transfer occurs during naturally more common implicit and unsupervised learning and if so, how it is related to memory consolidation. We compared the transfer of newly acquired explicit and implicit abstract knowledge during unsupervised learning by extending a visual statistical learning paradigm to a transfer learning context. We found transfer during unsupervised learning but with important differences depending on the explicitness/implicitness of the acquired knowledge. Observers acquiring explicit knowledge during initial learning could transfer the learned structures immediately. In contrast, observers with the same amount but implicit knowledge showed the opposite effect, a structural interference during transfer. However, with sleep between the learning phases, implicit observers switched their behaviour and showed the same pattern of transfer as explicit observers did while still remaining implicit. This effect was specific to sleep and not found after non-sleep consolidation. Our results highlight similarities and differences between explicit and implicit learning while acquiring generalizable higher-level knowledge and relying on consolidation for restructuring internal representations.

## Introduction

Recent years have seen a rise in interest in *transfer learning*, the application of previously learned regularities to novel input, as the effectiveness of this process is a key feature for biological and artificial intelligence alike (*1–3*). An important factor for successful transfer is the formation of abstract, generalizable knowledge transcending simple representation of the specific surface properties of the input (*4–7*). A challenging aspect of this memory formation is balancing the search for new commonalities between different tasks or contexts and minimizing the interference between new experiences and old memories (*8–10*). Previous empirical studies on transfer learning focused on the domains of supervised or reinforcement learning and mainly dealt with explicit knowledge. Although this allows for straightforward designs and ensures continuity with previous research, it ignores the fact that a large part of ecologically relevant learning happens implicitly (*11*) and in the absence of supervision or reinforcement (*12*, *13*). It is, therefore, an open question whether and how humans' transfer learning operates during unsupervised learning and in the absence of explicit knowledge.

In the current study, we contrast implicit and explicit transfer learning in an unsupervised learning paradigm. Directly contrasting implicit and explicit learning is important, as previous studies demonstrated that explicit and implicit forms of learning can show different outcomes and learning trajectories (*14–19*). In addition, we investigated the effect of consolidation on implicit transfer learning at various timescales since the role of consolidation in structural abstraction and transfer has been demonstrated previously for explicit learning with supervision or feedback (*20–25*) and since interactions between explicitness and consolidation exist (*26–30*). Our paradigm builds on the classical spatial visual statistical learning (SVSL) design (*31*) and extends it to a transfer learning setup. In this new setup, participants proceed from learning reappearing patterns embedded in unsegmented input to abstracting their shared structure and finally applying this structure to novel unsegmented input. This paradigm, therefore, combines the challenge of segmentation based on co-occurrence statistics (statistical learning) with the abstraction and

reapplication of shared structures (transfer learning), setting our study apart from previous studies on implicit abstraction or generalization (*32–34*).

Our results show that during unsupervised learning, the transfer of learned structures is immediately possible when explicit knowledge is acquired. In contrast, when only implicit knowledge is obtained, immediate transfer leads to an opposite effect, an interference between old knowledge and new learning. However, after 12-hour sleep consolidation following the learning of the first structure, while the knowledge still remains implicit, the same successful transfer of learned structures (i.e., generalization) occurs as in the explicit condition. Hence, our results extend the empirical findings on transfer learning to the domain of unsupervised implicit learning by showing that transfer learning is possible based on both explicit and implicit knowledge, albeit with different trajectories and highlight two important causes for these different trajectories: explicitness and consolidation of the knowledge to be transferred.

## Results

### *Visual Statistical Learning as a Testbed for Unsupervised Hierarchical Structure Learning*

Our experimental setup builds on the standard spatial visual statistical learning framework (SVSL)(*31*) and extends it to a transfer learning paradigm (Figure 1). In the traditional SVSL, participants passively watch a stream of scenes, each consisting of multiple shapes on a grid layout, which, unbeknown to the participant, are composed of a set of fixed shape pairs. The pairs are “fixed” since the two constituent shapes of a pair always appear next to each other in the scenes in a fixed spatial relationship. Importantly, each pair abuts at least one other pair in the scenes, so the pairs are not separated visually from each other. Thus, the identities of the shape pairs need to be extracted by learning across multiple scenes to achieve successful segmentation. Adult and infant humans, as well as various other species,

have been shown to automatically and implicitly extract the underlying pair structure of such scenes (*35–37*).

This paradigm is an ideal candidate for investigating unsupervised learning of higher-order structures as it not only allows learning the fixed pairs of shapes (chunks), but through these chunks, more abstract underlying structures can also be defined. One example of such a more abstract feature is the mean orientation of the learned chunks, where the orientation of a chunk is defined by whether the arrangement of the shapes within the pair is horizontal or vertical. Critically, these more abstract features are not observable properties of the scenes *per se,* as there are no segmentation cues revealing the chunks instantaneously. Thus, only when the chunks are learned can their orientation emerge as a feature. Hence, the latent property of orientation can be extracted only in conjunction with learning the observable pair associations. We used this feature in our transfer learning version of the SVSL paradigm, but we went beyond simply assessing whether a previously presented latent structure, such as the mean orientation of the learned chunks, became accessible to the observers. Instead, we measured how exposure to one type of abstract structure differentially influenced the acquisition of multiple types of structures later on.

Specifically, in the first training phase (Phase 1) in all of our experiments conducted online (Methods), participants saw scenes composed from only horizontal pairs or only vertical pairs, depending on the assigned condition counterbalanced across participants. In a second training phase (Phase 2), following some delay, they saw scenes composed from both horizontal and vertical pairs and made from a novel set of shapes not used in Phase 1. In the test phase after Phase 2, participants completed 2-alternative forced-choice trials (2AFC) where in each trial, they indicated whether they found more familiar a presented real pair (used in the training phases) or a foil pair (generated by randomly combining two shapes used in the training phase) (Figure 1). Our main question of interest was how unsupervised learning instances of only one latent structure (e.g., horizontal orientation of all chunks) in Phase 1 influenced the unsupervised learning of instances of the same

(horizontal) and a novel (vertical) structure in Phase 2. Therefore, our main measure of interest is the difference in learning the same vs. different types of pairs during Phase 2.

The explicitness of the acquired knowledge was assessed by using an exit survey at the very end of the experiment. Participants were labeled as “explicit” if they gave any indication of being aware of reappearing fixed patterns/pairs in the scenes. Additionally, a small number of participants who clearly reported the underlying structure (horizontality/verticality) were excluded from the analysis. Therefore, the participants we label “explicit” also had only far-from-perfect, partial knowledge about the structure of the scenes. In agreement with earlier visual statistical learning reports, the proportion of the excluded participants was 1-1.5% across experiments and their exclusion has not changed any of our results.

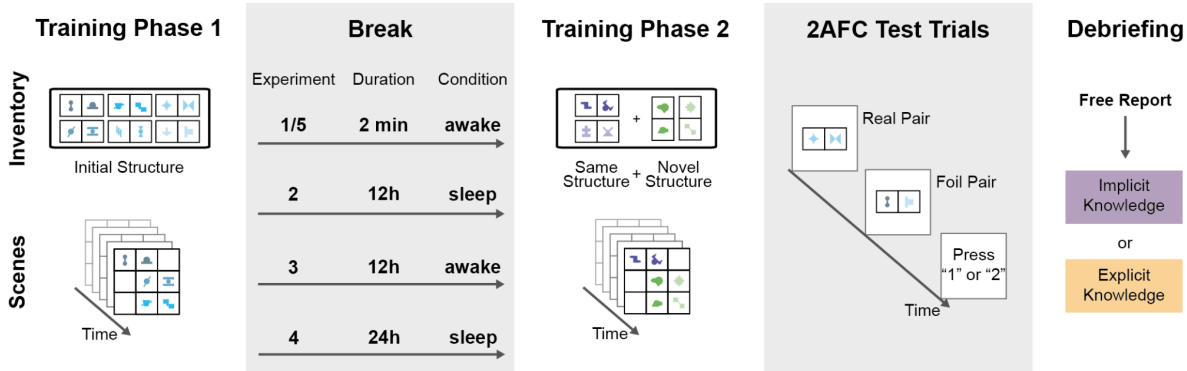



**Figure 1.** Overview of the experimental setup. **Training Phase 1:** Participants passively observed a stream of scenes made up of abstract shapes *(lower panel)*. Unbeknown to the participants, shapes in the scene appeared only in pairs of fixed spatial configurations defined by the Inventory *(upper panel)*. All pairs in Phase 1 had the same underlying structure of either horizontal or vertical orientation. The colors in the figure are only for illustration purposes; for the participants, all shapes were black. **Break:** After Phase 1, there was a break the length of which varied across the five experiments between two minutes and 24 hours. Participants spent the break either in asleep or awake condition. **Training Phase 2:** After the break, the participants were exposed to visual scenes made of a different set of abstract shapes. Half of the created pairs of the new inventory had horizontal, while the other half had vertical underlying structures. **2AFC Test Trials:** After Phase 2, participants completed a series of *2AFC Test Trials,* in which they had to decide if a real pair from the training phases or a foil pair, created by random combination of the shapes, was more familiar. **Debriefing:** Finally, participants answered open-ended questions about the experiment, which were used to assess whether they gained explicit knowledge about the presence of shape pairs.

***Explicit Learners Generalize while Implicit Learners Show a Structural Novelty Effect***

Experiment 1 (n=226 after exclusions; see Methods for details) investigated the immediate transfer of knowledge about the higher level structure between the two learning contexts by implementing a short 2-minute break between Phase 1 and 2 and thereby providing a baseline for the subsequent experiments. Both explicit and implicit learners in Experiment 1 performed above chance for pairs of the first training phase (Exp: M=67.9, SE=4.6, d=0.67, t(33)=3.93 p=0.002, BF=70.8; Impl: M=55.0, SE=1.1, d=0.32, t(191)=4.46 p<0.001, BF=919). In addition, the explicit participants (n=34) performed significantly better than the implicit ones (n=192) (d=0.73, t(224)=2.75 p=0.009, BF=195) (Figure 2). In Phase 2, explicit participants performed above chance for learning pairs that shared their higher level orientation structure with that of pairs in Phase 1 (M=67.6, SE=6.3, d=0.48, t(33)=2.81 p=0.033, BF=5). The same participants showed some moderate learning of the new pairs with the novel (non-matching) structure as well, but this learning failed to reach significance (M=58.1, SE=5.6, d=0.25, t(33)=1.46 p=0.465, BF=0.48). In contrast to the explicit participants, implicit participants showed the opposite pattern, performing significantly above chance for pairs of the novel structure (M=57.8, SE=1.9, d=0.30, t(191)=4.16 p<0.001, BF=281) but demonstrated strong evidence that they did not learn pairs sharing the higher level orientation structure with pairs in Phase 1 (M=49.1, SE=2.1, d=0.03, t(191)=-0.44 p=0.659, BF=0.09). The difference between the implicit participants' performances with same vs. different higher-order structures was statistically significant ($M_{diff}$=8.72, d=0.20, t(191)=2.78 p=0.030, BF=3.38). This qualitative pattern of the implicit participants' markedly better performance with novel structures was distinctively different from the pattern shown by the explicit participants, as indicated by the significant interaction between the two factors (participant type and structure type) of the two-factor mixed ANOVA results (F(1,224)=4.89, p=0.028, BF=31.6, $\eta_p^2$=0.02). These results imply that participants with more explicit acquired knowledge in Phase 1 effectively generalized the higher-level structure of this learned knowledge to novel situations and contexts, showing a "structure transfer" effect.

Meanwhile, implicit participants showed a “structural novelty” effect that might be explained by a structure-level interference due to a larger representational overlap between previously learned and presently seen horizontal pairs than between previous horizontal and present vertical pairs.

To strengthen the reliability of our results, we conducted a conceptual replication of Experiment 1 using two types of diagonal pairs, which were again orthogonal to each other. The results again showed a structural novelty effect for implicit participants (see Supplementary Experiment 1).

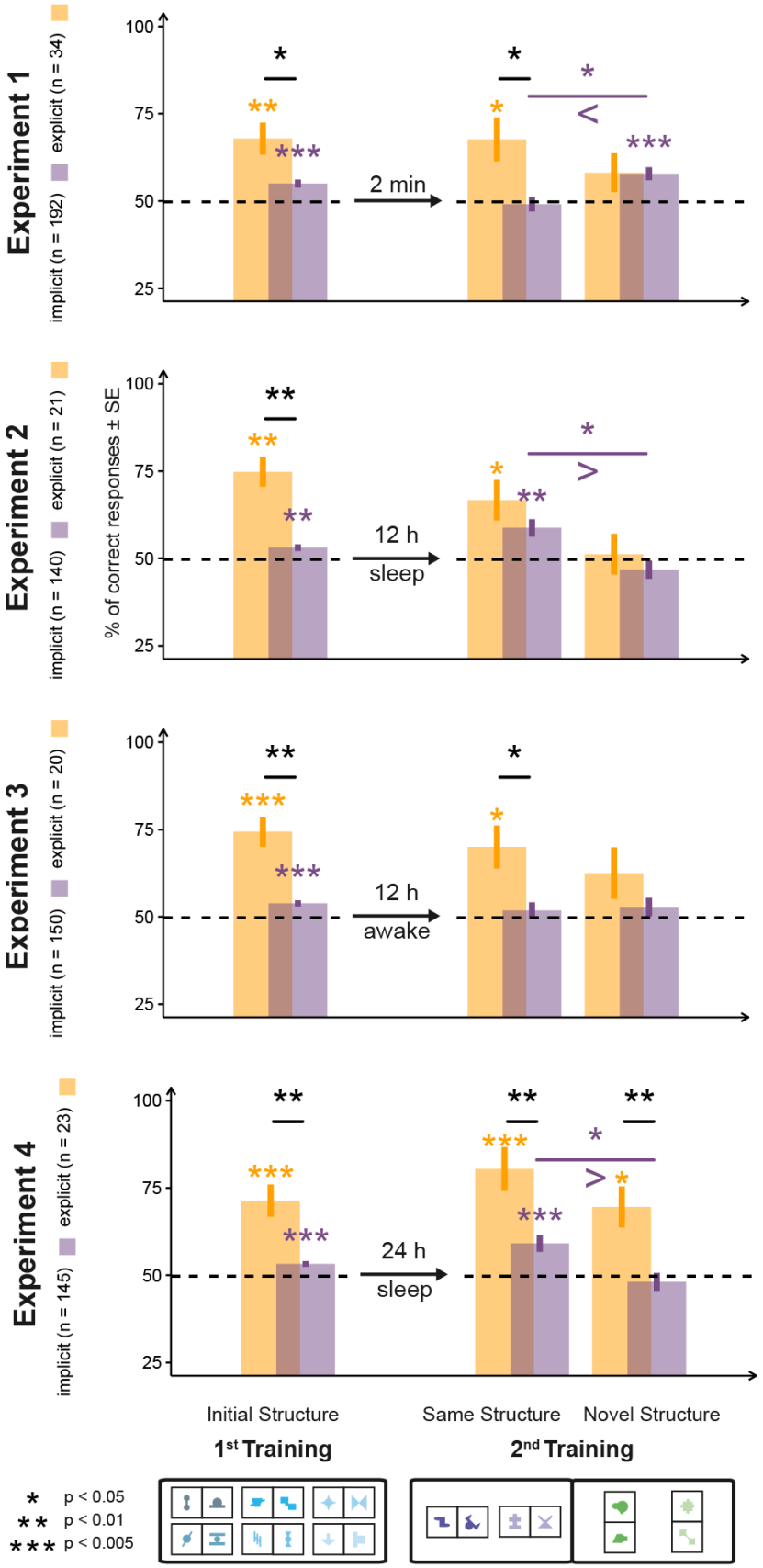



**Figure 2** Results of familiarity tests in Experiments 1, 2, 3, and 4. Test results of the 2AFC trials in all three experiments are grouped on the x-axis according to whether

the trials used shapes of the 1st or the 2nd training and, within the 2nd training, whether the pair in the trial had the same or different orientation structure as inventory pairs in the 1st training. The y-axis represents the proportion of correct responses in the 2AFC test trials. Arrows and text between the test results related to the two trainings convey the condition and length of the break period. Bars represent SEM, color coding indicates implicit and explicit subgroups of the participants. The horizontal dotted line denotes chance performance. Asterisks above bars denote significance levels from chance, while above lines, significance level comparing two conditions below the tips of the line. Legend of significance levels is shown in the lower left corner. Signs of inequality below the comparison in the 2nd training indicate the direction of effect.

***Consolidation Enables Implicit Learners to Generalize***

To investigate the effect of consolidation on explicit and implicit learners' structural transfer in Experiment 2 (n=161 after exclusions; see methods for details), we closely followed the design of Experiment 1 but introduced a 12-hour overnight consolidation phase between Phases 1 and 2. We found that while the performance of the explicit learners (n=21) after overnight sleep consolidation did not change drastically, it completely altered the pattern of behaviour of implicit learners (n=140) as they demonstrated in this condition the same generalization as the explicit learners did in Experiment 1 (Figure 2).

Specifically, explicit learners (n=21) performed above chance for pairs of the same structure (M=66.7, SE=5.8, d=0.63, t(20)=2.87 p=0.038, BF=5.3) but not pairs of a novel structure (M=51.2, SE=5.8, d=0.04, t(20)=0.20 p=0.841, BF=0.23) in in the tests of Phase 2 pairs. Implicit learners also performed above chance for pairs of the same structure (M=58.8, SE=2.5, d=0.30, t(139)=3.51 p=0.004, BF=37.6), but not pairs of the novel structure (M=46.8, SE=2.6, d=0.11, t(139)=-1.24 p=0.435, BF=0.27). The performance difference between these two types of pairs was significant ($M_{diff}$=11.96, d=0.24, t(139)=2.82 p=0.027, BF=5.4). A direct comparison of implicit participants' performance in Experiments 1 and 2 showed that the participants performed higher after sleep for same structure pairs and lower for novel structure pairs (see next sections for detailed results). Overall, the results of Experiments 1 and 2 demonstrate that while explicit participants could immediately

generalize the structure they learned, implicit participants required a consolidation period before being able to do the same.

### *The Effect of Consolidation is Specific to Sleep*

To clarify whether the effect observed in Experiment 2 is specific to sleep or just a general effect of consolidation, in Experiment 3 (n=170 after exclusions; see methods for details), we used the same general procedure as in Experiment 2 but with the 12-hour consolidation phase occurring during the day. The results of the implicit learners (n=150) provided strong evidence against generalization, as they performed at chance with both pairs of the same structure (M=51.8, SE=2.3, d=0.07, t(149)=0.80 p=0.999, BF=0.17) and pairs of a novel structure (M=52.8, SE=2.6, d=0.09, t(149)=1.08 p=0.999, BF=0.22) of Phase 2 (Figure 2). In contrast, participants with explicit knowledge (n=20) replicated the results of Experiments 1 and 2, with percent correct above chance for pairs of the same structure (M=70.0, SE=6.2, d=0.72, t(19)=3.24 p=0.026, BF=10.4) but not for pairs of a novel structure (M=62.5, SE=7.4, d=0.38, t(19)=1.70 p=0.530, BF=0.78).

### *The Effect of Sleep is not Explained by a Time-of-Day Effect*

It has previously been suggested that an apparent effect of sleep shown in an AM-PM vs PM-AM design can be based on the time of day at testing rather than sleep itself(*38*). To control for this potential confound, Experiment 4 (n=168 after exclusions; see methods for details) replicated Experiment 2 but with the second session delayed by 24 instead of 12 hours. Participants in this condition, therefore, had overnight sleep but were tested at the same time of day as the non-sleeping participants in Experiment 3. The results of Experiment 4 replicated the results of Experiment 2 for implicit learners, showing stronger learning for pairs of the same structure as compared to pairs of the novel structure

($M_{diff}$=11.04, d=0.23, t(144)=2.75 p=0.021, BF=4.4). We, therefore, conclude that the difference for implicit learners between our previous sleep and non-sleep conditions was not based on a time-of-day effect. For explicit learners, Experiment 4 showed no significant difference between learning for pairs of the same structure as compared to pairs of the novel structure ($M_{diff}$=10.78, d=0.26, t(22)=1.27 p=0.437, BF=0.44) and no significant correlation between learning pairs of the first learning phase and pairs of the novel structure (r=0.109, p=0.620). This suggests that the generalization effect found for explicit learners in the previous experiments is weakened during the longer consolidation phase of Experiment 4.

To compare directly the differential effect of type of consolidation on implicit structure learning, we entered the data of participants with implicit knowledge from Experiments 1, 2, 3, and 4 into a 4x2 ANOVA, with consolidation type (no consolidation, 12-h-sleep, 12-h-awake, and 24-h-sleep consolidation) and pair type (same or novel structure) as factors. The obtained results showed the typical pattern of a cross-over interaction with no significant main effects (consolidation type: F(3,623)=0.18, p=0.910, BF=0.003, $\eta_p^2$=0.0009; pair type: F(1,623)=1.52, p=0.218, BF=0.17, $\eta_p^2$=0.002) but a significant interaction (F(3,623)=7.43, p<0.001, BF=1979, $\eta_p^2$=0.03). Post-hoc tests revealed significant differences between the no-consolidation group (Exp. 1) and the two asleep-consolidation groups (Exp. 3 and 4), where the no-consolidation group showed stronger learning of novel structure pairs (Exp. 1 vs. Exp. 2: p=0.004, BF=44.3; Exp. 1 vs. Exp. 4: p=0.012; BF=12.7), while the asleep-consolidation groups showed stronger learning of same structure pairs (Exp. 1 vs. Exp. 2: p=0.015, BF=8.8; Exp. 1 vs. Exp. 4: p=0.011; BF=14.3). No other significant differences were found. These results confirm for implicit learners a directly opposite pattern of generalization behaviour between no-consolidation and asleep-consolidation conditions.

Additionally, to test for the presence of potential time-of-day effects in Experiment 1, we reanalyzed the data of Experiment 1 by taking into account the time point of testing. Both correlational and subgroup analyses found no indication of an effect of time of day on the pattern of structural transfer (see Supplementary Material).

***The Type of Transfer Depends on Quality of Knowledge, Not Quantity of Knowledge***

The different patterns of structural transfer for explicit and implicit participants could be based on either the quality of knowledge, i.e., its explicitness, or the quantity of knowledge, i.e., how much was learned during the first training phase. However, these two factors are confounded since the explicit participants typically performed higher for pairs of the first training phase. To address this confound, we conducted a matched sample analysis(*39*) in Experiments 1, 2, 3, and 4 to clarify which aspect of knowledge was responsible for our results. We selected a subsample of our implicit participants so that their accuracy performance matched that of the explicit participants for the first training phase and performed the same analyses on these subsampled populations as on the entire dataset in Experiments 1, 2, 3, and 4 (for details see Supplementary Materials). Matched implicit participants showed the same overall pattern of generalization behaviour as the full sample of implicit participants for all four experiments (Figure 3), although this failed to reach significance for Experiment 4. Specifically, participants learned more pairs of the novel than of the same structure in Experiment 1 ($p=0.012$; $BF=3.6$), they learned more pairs of the same than of the novel structure in Experiment 2 ($p<0.001$, $BF=127$), and they showed no significant difference between learning the two types of pairs in Experiment 3 ($p=0.214$, $BF=0.59$) and Experiment 4 ($p=0.304$, $BF=0.46$). With the matched sample analysis, we drastically reduced the sample size of the implicit participants, therefore reducing the power to detect the small effects found for this group. However, we recovered the same descriptive pattern in the full data set and the matched sample groups in all experiments and failed to recover the same statistical significance only in the time-of-the-day control experiment. Therefore, we posit that these findings strongly support the notion that the difference in the structural transfer is based predominantly on the quality of knowledge - its explicitness - not the exact level of learning.

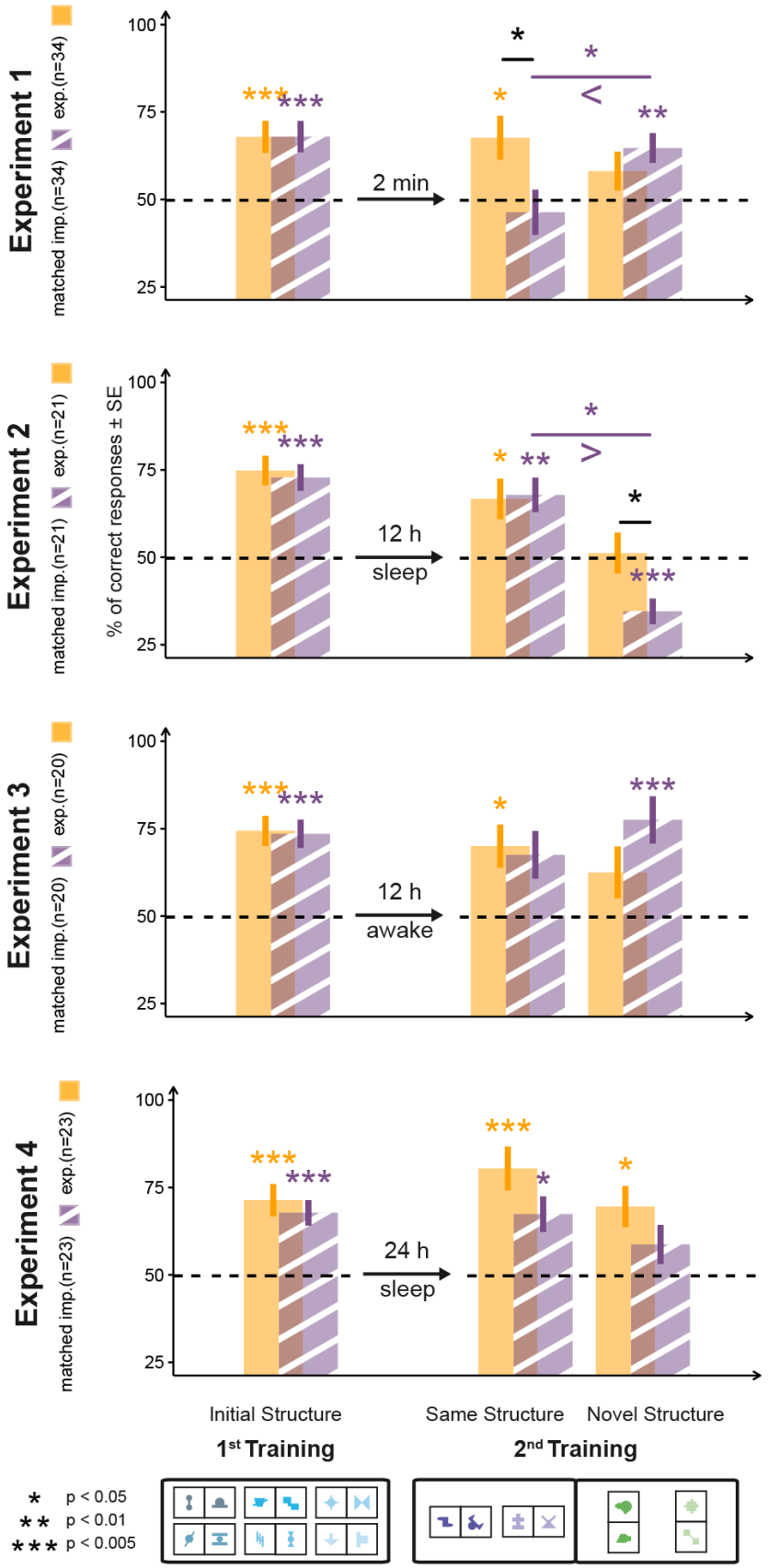



**Figure 3**. Matched Sample Analysis of Experiments 1, 2, 3, and 4. The structure of the figure is identical to that of Figures 1 and 2, with data for the explicit participants (orange bars)

being the same as in Figure 1, while data for the implicit participants (striped purple bars) showing the subgroup of implicit participants from Experiment 1 whose combined performance matched the performance of the explicit participants on test trials of the first learning phase. The y-axes represent the proportion of correct responses in the 2AFC trials. Bars represent the mean (±SEM) for each type of pair (pairs of Phase 1 and same and novel structure pairs of Phase 2). The horizontal dashed line indicates chance performance.

***Explicitness and Generalization can be Induced by Verbal Instruction***

The explicit-implicit parameterization in Experiment 1 was quasi-experimental rather than true experimental since the groups were formed naturally. Therefore, it is unclear whether it is indeed explicitness that enables generalization or whether the two groups of participants are different in other important ways (e.g., task engagement, attentional processes). To investigate this issue, in Experiment 5 (n=36 after exclusions) we rerun Experiment 1 with a new group of participants and with the single change in the protocol that explicitness of knowledge was induced via explicit verbal instructions stating that all shapes would be grouped into pairs. The results of Experiment 5 closely matched those of the explicit participants of Experiment 1. The participants had an above chance performance for Phase 1 ($M=71.8$, $SE=3.6$, $d=1.01$, $t(35)=6.07$ $p<0.001$, $BF=26395$) and for the same structure ($M=72.9$, $SE=4.0$, $d=0.95$, $t(35)=5.69$ $p<0.001$, $BF=9058$) but not novel structure pairs ($M=52.8$, $SE=5.8$, $d=0.08$, $t(35)=0.48$ $p=0.634$, $BF=0.20$) in Phase 2. The performance for pairs of the same and novel structure was significantly different ($d=0.64$, $t(35)=3.83$ $p=0.001$, $BF=58.0$). These results confirm that the generalization producing an immediate transfer between contexts can be easily induced by verbal instructions in any population and therefore, the type of explicitness studied in our experiments is a likely candidate of the necessary condition for such a generalization.

In Experiments 1-4 and Supplementary Experiment 1 explicit participants showed the same descriptive pattern of higher performance for same over novel structure pairs, but this difference failed to reach significance within experiments. Therefore, we reanalyzed these results by collapsing the data over all these experiments. The results showed an overall

higher performance on same structure pairs than novel structure pairs for explicit participants ($d=0.34$, $t(233.99)=2.56$, $p=0.011$, BF=3.12). Furthermore, over all experiments, explicit participants' performance for Phase 1 only significantly correlated with learning of same structure pairs ($r=0.41$, $p<0.001$, BF=4460), not novel structure pairs ($r=0.14$, $p=0.137$, BF=0.62).

## Discussion

Our findings demonstrated for the first time an interaction of consolidation and explicitness of knowledge in humans' unsupervised transfer learning. Participants with explicit knowledge are able to immediately transfer structural knowledge from one learning context to another. In contrast, participants with implicit knowledge show a structural novelty effect in immediate transfer. Only after a phase of asleep consolidation are they able to generalize from one learning context to another.

Our interpretation of these results is that asleep consolidation leads to an internal redescription of the acquired knowledge by factorization, that is, by representing the feature of orientation as a useful summary statistic in its own right. This redescription of the input with a hierarchical model using a higher-level representation of the underlying orientation structure emerges as a complement to the chunk representation learned by statistical learning during exposure. While without such a redescription, the previously learned patterns in Phase 1 interfere with new ones in Phase 2, when redescription occurs, the abstract knowledge at the higher level of the hierarchy can be generalized i.e., it can be used during the processing and segmentation of new input. Instead of a simple feed-forward hierarchy progressing from specific to abstract, this interpretation implies a hierarchical system with bi-directional interactions: learning representations at a higher level of abstraction (orientation) constrain the learning of new representations of lower-level features (actual chunks) later in time. Thus our approach to hierarchy is different from earlier treatments of

SL in hierarchical systems that focused on hierarchies of composition (*36*) or hierarchies that arbitrated between competing learning systems (*40*).

One notable consequence of the emergence of such hierarchies is linking SL more to the emergence of object representations. Previous studies demonstrated that statistical learning is not based solely on momentarily observable co-occurrence statistics but is also biased by prior knowledge. However, these studies demonstrate various effects of prior knowledge acquired over a lifetime (*41*) or shorter period (*42–44*), provided explicitly (*45*), or facilitated by segmentation cues (*46*). They also focused mostly on the language domain or sequential streams and either relied on an uncontrolled amount of long-term knowledge or investigated token-level effects. In contrast, the current study investigated structure-related effects that could not be explained by the transfer of token-level knowledge, and it strictly controlled the amount of previous exposure to the target structure. Our results demonstrate that similar to objects (*47*, *48*), chunks acquired during SL are not represented simply as an unrelated set of inventory elements but organized into higher-level categories based on the common underlying structures - e.g., the category of horizontal pairs. Thus, our results build a bridge between classical SL and category learning, and, in line with earlier findings (*49*, *50*), it supports the idea that SL is a vehicle for acquiring object-like representations.

Previous studies on statistical learning and consolidation focused on how much time or sleep helps with the stabilization or improvement of specific memory traces and obtained mixed results, with some studies showing clear effects (*20*, *51–53*) and others showing no or very limited effects of sleep (*54–60*). In addition, most of these studies did not consider the actual state of the acquired knowledge (but see (*29*)). In contrast, our study specifically focused on the difference between explicit and implicit learners to disentangle the complex nature of consolidation and we use the ability to generalize structure between learning contexts as our primary measure. This approach allowed us to compare our findings to those reported in the domain of explicit learning (*20*, *21*, *24*, *61*) and uncover a surprising similarity between consolidation benefits based on abstraction and generalization in the two domains.

In particular, similarly to observers who needed sleep to “discover” abstract learning rules in an explicit task (*61*), our implicit learners required sleep to “uncover” abstract higher-level descriptors of the observed structures and utilize them in subsequent implicit learning. This parallel in behaviour suggests an analogous process for integrating and using implicitly and explicitly acquired new information at different levels of the existing internal representation.

Previous statistical learning experiments reported that quantitative performance improves by explicit instructions (*45*) and that performances in the presence and absence of explicit instructions converge over longer exposure (*62*). Our results in Experiment 5 are in line with those findings: they show stronger performance following explicit instructions about the task structure and convergence between explicit and implicit learning. In addition, it goes beyond these findings by demonstrating that vanishing of differences after sleep consolidation occurs even when prior to consolidation there is a qualitative rather than only quantitative difference between the initial implicit and explicit learning results.

In summary, our work builds multiple bridges between different domains of cognition. First, it relates low-level correlational learning and higher-level structure learning in an unsupervised transfer learning setup that is separate from the classical problem- or task-solving paradigms (*26*, *63*, *64*) yet can be readily interpreted in those frameworks. This allows extending the definition of higher-level structure to a domain where explicit tasks and thought are not defined. Second, it reports a new repulsion-attraction effect at the level of implicit structure transfer, namely a tradeoff between generalization and interference depending on the presence or absence of sleep-consolidation processes. Previously, such repulsion-attraction effects have been found only within immediate perceptual processes without being related to sleep, such as implicit low-level perception (*65*–*67*) or high-level categorization (*20*, *21*, *24*, *61*, *68*) involving sleep but requiring explicit tasks. These converging findings might indicate the existence of a more general factorization mechanism in the brain that works continuously and at multiple levels during perceptual and cognitive processes following the same computational principles. However, confirming this conjecture

will require further investigations of the interplay between learning and consolidation in humans. Application of our approach could help to resolve this issue by facilitating an integrated exploration of how the brain learns complex internal representations and utilizes them in a recursive manner.

**Data Availability**

The experimental data that support the findings of this study are available on OSF: https://osf.io/untbz

## Methods

### *Experiment 1*

*Participants*

251 participants (92 female, Age: mean = 28.0, mode = 25, SD = 9.5) were recruited via prolific.co. All participants had normal or corrected-to-normal vision and provided informed consent.

*Materials*

The stimuli were taken from(*31*) and consisted of 20 abstract black shapes on white background (see Figure 1). The shapes were grouped to form six pairs of the same orientation (horizontal or vertical) for the first learning phase and four pairs, two horizontal and two vertical, for the second learning phase. The assignment of shapes to pairs was randomized for each participant. Scenes were created by placing three pairs together on a 3x3 grid without segmentation cues. 160 scenes were created for the first and 48 for the second learning phase. In the second learning phase, each scene was used twice for a total of 96 presented scenes. The study was conducted online via prolific (see Supplementary Materials for details).

*Procedure*

Participants passively observed 160 scenes in the first training phase. For half of the participants, these scenes contained only horizontal pairs (horizontal condition), and for the other half only vertical pairs (vertical condition). Each scene was presented for 2 s with a 1 s interstimulus interval (ISI). After a two-minute passive break, participants passively observed 96 scenes in the second training phase. Participants were not told about the presence of any structure in the scenes. Pair learning was tested with a two-alternative forced choice task (2AFC). In each trial, participants saw a real shape pair from one of the training phases and

a foil pair created by combining shapes from two different pairs of the same training phase. Participants were asked to indicate which of the two was more familiar by pressing “1” or “2” on their keyboard. Finally, participants answered five open questions about their beliefs about the experiment and their knowledge of pair structure. For details see Supplementary Materials.

*Results*

229 participants remained after exclusions (see Supplementary for details). Based on the open responses at the end of the experiment participants were categorized into one of three groups. Participants who reported no knowledge of pairs were counted as implicit (n=192), participants who reported knowledge of the presence of pairs were counted as explicit (n=34), and participants who also reported the underlying horizontal/vertical structure were excluded from analysis as they were too few for meaningful analysis (n=3).

Test trials for the second training phase were scored separately for pairs of the same structure as in the first training phase and pairs of the novel structure. Explorative analysis of the trials of the second training phase revealed a strong negative correlation between trials of the novel and the same structure ($r=-0.432$, $p<0.001$). As the foil pairs used for one type of structure were created by recombining shapes of pairs of the other structure, this could be a type of consistency effect where participants tend to choose the same shapes, independent of pair knowledge. If this was the case we expect to see an increase in this negative correlation as participants complete more test trials. Indeed, for the first 8 trials, this correlation is $r=-0.209$. while it increases to $r=-0.435$ in the second 8 trials. Based on this exploration we included only the first 8 trials in all the following analyses to minimize this consistency effect. To keep comparability, the following experiments use the same number of test trials but we again only analyze the first 8. The data was collapsed over vertical and horizontal conditions for all analyses (see Supplementary Materials).

The results (Figure 2) show above chance performance for the first training phase for both explicit (M=67.9, SE=4.6, d=0.67, t(33)=3.93 p=0.002, BF=70.8) and implicit (M=55.0, SE=1.1, d=0.32, t(191)=4.46 p<0.001, BF=919) participants. However, the performance was significantly higher for explicit participants (d=0.73, t(224)=2.75 p=0.009, BF=195). For the second learning phase, implicit participants show above chance performance for pairs of the novel structure (M=57.8, SE=1.9, d=0.30, t(191)=4.16 p<0.001, BF=281), and chance performance for pairs of the same structure as before (M=49.1, SE=2.1, d=0.03, t(191)=-0.44 p=0.659, BF=0.09). The performance for these types of pairs is significantly different from each other (d=0.20, t(382)=2.78 p=0.030, BF=3.38). Explicit participants show the opposite pattern with above-chance performance for pairs of the same structure as before (M=67.6, SE=6.3, d=0.48, t(33)=2.81 p=0.033, BF=5.0), and chance performance for pairs of the novel structure (M=58.1, SE=5.6, d=0.25, t(33)=1.46 p=0.465, BF=0.48). However, for them, this difference is not significant (d=0.19, t(33)=1.12 p=0.539, BF=0.33). We do however see a significant, medium to large correlation between learning in the first training phase and learning pairs of the same structure in the second training phase (r=0.45, p=0.008) for the explicit participants. Such a correlation is not observed between learning in the first learning phase and learning pairs of the novel structure in the second learning phase (r=-0.01, p=0.947). The different patterns of transfer behaviour for explicit and implicit participants are also confirmed when entering the data into a 2x2 mixed ANOVA with factors *participant type* (explicit or implicit) and *structure type* (novel or same). The results show a significant main effect of *participant type* (F(1, 224)=8.03, p=0.005, BF=2.1, $\eta_p^2$=0.03)) and of *structure type* (F(1, 224)=4.05, p=0.045, BF=1.2, $\eta_p^2$=0.02)) as well as a significant interaction of both (F(1,224)=4.89, p=0.028, BF=31.6, $\eta_p^2$=0.02).

***Experiment 2***

*Participants*

259 participants (127 female, Age: mean = 25.6, mode = 21, SD = 8.5) were recruited via prolific.co. All participants had normal or corrected-to-normal vision and provided informed consent.

*Materials*

The materials of the main part of the experiment were identical to Experiment 1. Additionally, participants filled out the Pittsburgh Sleep Quality Index (PSQI)(*69*) and the Groningen Sleep Quality Scale (GSQS)(*70*).

*Procedure*

The procedure within the main tasks was identical to Experiment 1. However, in this experiment participants completed the first training phase in the evening at 09:00 pm, followed by the GSGS and PSQI questionnaires. 12 hours later in the morning at 09:00 am, they completed the second training phase followed by all the test trials and finally, they completed the GSGS again.

*Results*

161 participants remained after exclusions (see Supplementary for details). Participants were categorized as explicit or implicit in the same way as in Experiment 1. This led to 21 participants being categorized as explicit and 140 as implicit. Bayes Factors from Bayesian t-tests for implicit participants reported for Experiments 2, 3, and 4 used an r-scale parameter of 0.5 instead of the default $\sqrt{2}/2$, reflecting that Experiment 1 found small effect sizes for this group.

The results (Figure 2) showed that participants with implicit knowledge (n=140) performed above chance for pairs of the first training phase (M=53.1, SE=1.0, d=0.27, t(139)=3.24 p=0.009, BF=16.8) and for pairs of the same structure (M=58.8, SE=2.5, d=0.30, t(139)=3.51 p=0.004, BF=37.6) but not pairs of a novel structure (M=46.8, SE=2.6, d=0.11, t(139)=-1.24 p=0.435, BF=0.27) in the second training phase. The performance for pairs of the same and novel structure is significantly different (d=0.24, t(139)=2.82 p=0.027, BF=5.4). Participants with explicit knowledge (n=21) show the same pattern of results as they did in Experiment 1, performing above chance for pairs of the first training phase (M=74.8, SE=4.2, d=1.3, t(20)=5.9 p<0.001, BF=2361) and for pairs of the same structure (M=66.7, SE=5.8, d=0.63, t(20)=2.87 p=0.038, BF=5.3) but not pairs of a novel structure (M=51.2, SE=5.8, d=0.04, t(20)=0.20 p=0.841, BF=0.23) in the second training phase. As in Experiment 1, for explicit participants, the difference between pairs of the same and novel structure is not significant (d=0.39, t(20)=1.77 p=0.272, BF=0.86), but we do see a strong positive correlation between learning in the first learning phase and learning pairs of the same structure in the second learning phase (r=0.52, p=0.015). Such a correlation is again not observed between learning in the first learning phase and learning pairs of the novel structure in the second learning phase (r=0.27, p=0.232).

***Experiment 3***

*Participants*

275 participants (134 female, Age: mean = 28.9, mode = 24, SD = 9.8) were recruited via prolific.co. All participants had normal or corrected-to-normal vision and provided informed consent.

*Materials*

The materials were identical to those of Experiment 2.

*Procedure*

The procedure was the same as Experiment 2, with the difference that the first session took place in the morning at 09:00 am and the second in the evening at 09:00 pm. As there was no night of sleep between the first and second sessions, participants filled out the GSQS only once in this experiment.

*Results*

170 participants remained after exclusions (see Supplementary for details). Participants were categorized as explicit or implicit in the same way as in Experiment 1. This led to 20 participants being categorized as explicit and 150 as implicit. Bayes Factors from Bayesian t-tests for implicit participants reported for Experiments 2, 3, and 4 used an r-scale parameter of 0.5 instead of the default $\sqrt{2}/2$, reflecting that Experiment 1 found small effect sizes for this group.

The results (Figure 2) showed that participants with implicit knowledge (n=150) that did not sleep during the consolidation phase performed above chance for pairs of the first training phase (M=53.9, SE=0.9, d=0.36, t(149)=4.46 p<0.001, BF=1020), but not for pairs of the same structure (M=51.8, SE=2.3, d=0.07, t(149)=0.80 p=0.999, BF=0.17) or pairs of a novel structure (M=52.8, SE=2.6, d=0.09, t(149)=1.08 p=0.999, BF=0.22) in the second training phase. Participants with explicit knowledge (n=20) again perform as they did in Experiment 1 performing above chance for pairs of the first training phase (M=74.4, SE=4.3, d=1.27, t(19)=5.67 p<0.001, BF=1281) and for pairs of the same structure (M=70.0, SE=6.2, d=0.72, t(19)=3.24 p=0.026, BF=10.4) but not pairs of a novel structure (M=62.5, SE=7.4, d=0.38, t(19)=1.70 p=0.530, BF=0.78) in the second training phase. As in Experiment 1, for explicit participants, the difference between pairs of the same and novel structure is not significant (d=0.19, t(19)=0.86, p=0.999, BF=0.32), but again we do see a strong positive correlation between learning in the first learning phase and learning pairs of the same structure in the second learning phase (r=0.56, p=0.010). Such a correlation is again not

observed between learning in the first learning phase and learning pairs of the novel structure in the second learning phase (r=0.42, p=0.065).

***Experiment 4***

*Participants*

275 participants (129 female, Age: mean = 27.9, mode = 23, SD = 8.9) were recruited via prolific.co. All participants had normal or corrected-to-normal vision and provided informed consent.

*Materials*

The materials were identical to Experiment 2.

*Procedure*

The procedure was identical to Experiment 2. However, in this experiment participants completed the first session in the evening at 09:00 pm, and the second session 24 hours later again in the evening at 09:00 pm.

*Results*

169 participants remained after exclusions (see Supplementary for details). Participants were categorized as explicit or implicit in the same way as in Experiment 1. This led to 145 participants being categorized as implicit, 23 categorized as implicit, and 1 being excluded for being categorized as fully explicit. Bayes Factors from Bayesian t-tests for implicit participants reported for Experiments 2, 3, and 4 used an r-scale parameter of 0.5 instead of the default √2/2, reflecting that Experiment 1 found small effect sizes for this group.

The results (Figure 2) showed that participants with implicit knowledge (n=145) performed above chance for pairs of the first training phase (M=53.2, SE=0.8, d=0.34, t(144)=4.09 p=0.001, BF=260) and for pairs of the same structure (M=59.1, SE=2.4, d=0.31, t(144)=3.78 p=0.001, BF=89.4) but not pairs of a novel structure (M=48.1, SE=2.6, d=0.06, t(144)=-0.74 p=0.459, BF=0.17) in the second training phase. The performance for pairs of the same and novel structure is significantly different ($M_{diff}$=11.04, d=0.23, t(144)=2.75 p=0.021, BF=4.4). Participants with explicit knowledge (n=23) show the same pattern of results as they did in Experiment 1, performing above chance for pairs of the first training phase (M=71.4, SE=4.6, d=0.9, t(23)=4.6 p=0.001, BF=217), for pairs of the same structure (M=80.4, SE=6.3, d=1.01, t(23)=4.85 p=0.001, BF=347), and pairs of a novel structure (M=69.6, SE=5.9, d=0.69, t(23)=3.33 p=0.012, BF=13.6) in the second training phase. As in Experiment 1, for explicit participants, the difference between pairs of the same and novel structure is not significant ($M_{diff}$=10.78, d=0.26, t(22)=1.27 p=0.437, BF=0.44), but here we do not see a significant positive correlation between learning in the first learning phase and learning pairs of the same structure in the second learning phase (r=0.109, p=0.620).

***Experiment 5***

*Participants*

40 participants (18 female, Age: mean = 28.4, mode = 19, SD = 11.8) were recruited via prolific.co. All participants had normal or corrected-to-normal vision and provided informed consent.

*Materials*

The materials were identical to Experiment 1.

*Procedure*

The procedure was identical to Experiment 1, apart from the instructions. In this experiment, participants were told about the pair structure before the beginning of the experiment. Participants were told that whenever a specific shape appears in the grid a second specific shape will appear in a fixed position near it.

*Results*

36 participants remained after exclusions (see Supplementary for details). Overall, the results show the same pattern as the explicit participants in Experiment 1, with above-chance performance for pairs of the first training phase ($M=71.8$, $SE=3.6$, $d=1.01$, $t(35)=6.07$ $p<0.001$, $BF=26395$) and for pairs of the same structure ($M=72.9$, $SE=4.0$, $d=0.95$, $t(35)=5.69$ $p<0.001$, $BF=9058$) but not pairs of a novel structure ($M=52.8$, $SE=5.8$, $d=0.08$, $t(35)=0.48$ $p=0.634$, $BF=0.20$) in the second training phase. The performance for pairs of the same and novel structure is significantly different ($d=0.64$, $t(35)=3.83$ $p=0.001$, $BF=58.0$).

## References


1. J. L. Austerweil, S. Sanborn, T. L. Griffiths, Learning How to Generalize. *Cogn. Sci.* **43**, e12777 (2019).

2. S. Mark, R. Moran, T. Parr, S. W. Kennerley, T. E. J. Behrens, Transferring structural knowledge across cognitive maps in humans and models. *Nat. Commun.* **11**, 4783 (2020).

3. J. X. Wang, Meta-learning in natural and artificial intelligence. *Current Opinion in Behavioral Sciences* **38**, 90–95 (2021).

4. V. Samborska, J. L. Butler, M. E. Walton, T. E. J. Behrens, T. Akam, Complementary task representations in hippocampus and prefrontal cortex for generalizing the structure of problems. *Nat. Neurosci.* **25**, 1314–1326 (2022).

5. J. C. R. Whittington, T. H. Muller, S. Mark, G. Chen, C. Barry, N. Burgess, T. E. J. Behrens, The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* **183**, 1249–1263.e23 (2020).

6. W. Sun, M. Advani, N. Spruston, A. Saxe, J. E. Fitzgerald, Organizing memories for generalization in complementary learning systems. *Nat. Neurosci.* **26**, 1438–1448 (2023).

7. R. B. Dekker, F. Otto, C. Summerfield, Curriculum learning for human compositional generalization. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2205582119 (2022).

8. T. Flesch, J. Balaguer, R. Dekker, H. Nili, C. Summerfield, Comparing continual task learning in minds and machines. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E10313–E10322 (2018).

9. T. Flesch, A. Saxe, C. Summerfield, Continual task learning in natural and artificial agents. *Trends Neurosci.* **46**, 199–210 (2023).

10. T. Flesch, D. G. Nagy, A. M. Saxe, C. Summerfield, Modelling continual learning in humans with Hebbian context gating and exponentially decaying task signals Short title: A neural network model of human continual learning. (2022).

11. A. S. Reber, Implicit Learning and Tacit Knowledge. *J. Exp. Psychol. Gen.* **118**, 219–235 (1989).

12. G. E. Hinton, Learning to represent visual input. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 177–184 (2010).

13. G. Hinton, Where do features come from? *Cogn. Sci.* **38**, 1078–1101 (2014).

14. R. C. Mathews, R. R. Buss, W. B. Stanley, F. Blanchard-Fields, J. R. Cho, B. Druhan, Role of Implicit and Explicit Processes in Learning From Examples: A Synergistic Effect.

15. R. Dale, N. D. Duran, J. R. Morehead, Prediction during statistical learning, and implications for the implicit/explicit divide. *Adv. Cogn. Psychol.* **8**, 196–209 (2012).

16. F. Ball, I. Spuerck, T. Noesselt, Minimal interplay between explicit knowledge, dynamics of learning and temporal expectations in different, complex uni- and multisensory contexts. *Atten. Percept. Psychophys.* **83**, 2551–2573 (2021).

17. A. Bloch, D. Tamir, E. Vakil, G. Zeilig, Specific Deficit in Implicit Motor Sequence Learning following Spinal Cord Injury. *PLoS One* **11**, e0158396 (2016).

18. E. Poh, J. A. Taylor, Generalization via superposition: combined effects of mixed reference frame representations for explicit and implicit learning in a visuomotor adaptation task. *J. Neurophysiol.* **121**, 1953–1966 (2019).

19. M. Forano, R. Schween, J. A. Taylor, M. Hegele, D. W. Franklin, Direct and indirect cues can enable dual adaptation, but through different learning processes. *J. Neurophysiol.* **126**, 1490–1506 (2021).

20. P. A. Lewis, S. J. Durrant, Overlapping memory replay during sleep builds cognitive schemata. *Trends Cogn. Sci.* **15**, 343–351 (2011).

21. S. Diekelmann, J. Born, The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).

22. B. Rasch, J. Born, About Sleep's Role in Memory. *Physiol. Rev.* **93**, 681–766 (2013).

23. A. M. Chambers, The role of sleep in cognitive processing: focusing on memory consolidation. *Wiley Interdiscip. Rev. Cogn. Sci.* **8** (2017).

24. I. Lerner, M. Gluck, Sleep and the Extraction of Hidden Regularities: A Systematic Review and the Importance of Temporal Rules. *Sleep Med. Rev.* **47**, 39–50 (2019).

25. J. G. Klinzing, N. Niethard, J. Born, Mechanisms of systems memory consolidation during sleep. *Nat. Neurosci.* **22**, 1598–1610 (2019).

26. U. Wagner, S. Gais, H. Haider, R. Verleger, J. Born, Sleep inspires insight. *Nature* **427**, 352–355 (2004).

27. S. Fischer, S. Drosopoulos, J. Tsen, J. Born, Implicit learning--explicit knowing: a role for sleep in memory system interaction. *J. Cogn. Neurosci.* **18**, 311–319 (2006).

28. T. Zander, K. G. Volz, J. Born, S. Diekelmann, Sleep increases explicit solutions and reduces intuitive judgments of semantic coherence. *Learn. Mem.* **24**, 641–645 (2017).

29. H. Liu, T. A. Forest, K. Duncan, A. S. Finn, What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition* **236**, 105439 (2023).

30. E. M. Robertson, A. Pascual-Leone, D. Z. Press, Awareness Modifies the Skill-Learning Benefits of Sleep. *Curr. Biol.* **14**, 208–212 (2004).

31. J. Fiser, R. N. Aslin, Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychol. Sci.* **12**, 499–504 (2001).

32. G. F. Marcus, S. Vijayan, S. Bandi Rao, P. M. Vishton, Rule learning by seven-month-old infants. *Science* **283**, 77–80 (1999).

33. A. S. Reber, Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior* **6**, 855–863 (1967).

34. R. L. Gómez, Variability and detection of invariant structure. *Psychol. Sci.* **13**, 431–436 (2002).

35. J. Fiser, R. N. Aslin, Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences* **99**, 15822–15826 (2002).

36. A. L. F. Lee, Z. Liu, H. Lu, Parts beget parts: Bootstrapping hierarchical object representations through visual statistical learning. *Cognition* **209**, 104515 (2021).

37. C. Santolin, J. R. Saffran, Constraints on Statistical Learning Across Species. *Trends Cogn. Sci.* **22**, 52–63 (2018).

38. M. C. Tandoc, M. Bayda, C. Poskanzer, E. Cho, R. Cox, R. Stickgold, A. C. Schapiro, Examining the effects of time of day and sleep on generalization. *PLoS One* **16**, e0255423 (2021).

39. D. E. Ho, K. Imai, G. King, E. A. Stuart, Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Polit. Anal.* **15**, 199–236 (2007).

40. M. Maheu, F. Meyniel, S. Dehaene, Rational arbitration between statistics and rules in human sequence processing. *Nat Hum Behav* **6**, 1087–1103 (2022).

41. K. Stärk, E. Kidd, R. L. A. Frost, Close encounters of the word kind: Attested distributional information boosts statistical learning. *Lang. Learn.* **73**, 341–373 (2022).

42. A. Kóbor, K. Horváth, Z. Kardos, D. Nemeth, K. Janacsek, Perceiving structure in unstructured stimuli: Implicitly acquired prior knowledge impacts the processing of unpredictable transitional probabilities. *Cognition* **205**, 104413 (2020).

43. D. M. Antovich, K. Graf Estes, Statistical word segmentation: Anchoring learning across contexts. *Infancy* **28**, 257–276 (2023).

44. J. Chough, B. D. Zinszer, Deep in the Trenches: First language performance predicts primacy in statistical learning of two structures. *Proceedings of the Annual Meeting of the Cognitive Science Society* **44** (2022).

45. J. Arciuli, J. von K. Torkildsen, D. J. Stevens, I. C. Simpson, Statistical learning under incidental versus intentional conditions. *Front. Psychol.* **5**, 747 (2014).

46. J. L. Mueller, C. T. Cate, J. M. Toro, A Comparative Perspective on the Role of Acoustic Cues in Detecting Language Structure. *Top. Cogn. Sci.* **12**, 859–874 (2020).

47. F. G. Ashby, W. T. Maddox, Human Category Learning. *Annu. Rev. Psychol.* **56**, 149–178 (2005).

48. J. J. Richler, T. J. Palmeri, Visual category learning. *Wiley Interdiscip. Rev. Cogn. Sci.* **5**, 75–94 (2014).

49. G. Lengyel, G. Žalalytė, A. Pantelides, J. N. Ingram, J. Fiser, M. Lengyel, D. M. Wolpert, Unimodal statistical learning produces multimodal object-like representations. *Elife* **8**, e43942 (2019).

50. G. Lengyel, M. Nagy, J. Fiser, Statistically defined visual chunks engage object-based attention. *Nat. Commun.* **12**, 272 (2021).

51. S. J. Durrant, S. A. Cairney, P. A. Lewis, Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cereb. Cortex* **23**, 2467–2478 (2013).

52. S. J. Durrant, S. A. Cairney, P. A. Lewis, Cross-modal transfer of statistical information benefits from sleep. *Cortex* **78**, 85–99 (2016).

53. N. Hennies, P. A. Lewis, S. J. Durrant, J. N. Cousins, M. A. L. Ralph, Time- but not sleep-dependent consolidation promotes the emergence of cross-modal conceptual representations. *Neuropsychologia* **63**, 116–123 (2014).

54. R. Kim, A. Seitz, H. Feenstra, L. Shams, Testing assumptions of statistical learning: Is it long-term and implicit? *Neurosci. Lett.* **461**, 145–149 (2009).

55. D. Nemeth, K. Janacsek, Z. Londe, M. T. Ullman, D. V. Howard, J. H. Howard Jr, Sleep has no critical role in implicit motor sequence learning in young and old adults. *Exp. Brain Res.* **201**, 351–358 (2010).

56. J. Arciuli, I. C. Simpson, Statistical learning is lasting and consistent over time. *Neurosci. Lett.* **517**, 133–135 (2012).

57. E. Hallgató, D. Győri-Dani, J. Pekár, K. Janacsek, D. Nemeth, The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex* **49**, 1073–1081 (2013).

58. P. Simor, Z. Zavecz, K. Horváth, N. Éltető, C. Török, O. Pesthy, F. Gombos, K. Janacsek, D. Nemeth, Deconstructing Procedural Memory: Different Learning Trajectories and Consolidation of Sequence and Statistical Learning. *Front. Psychol.* **9**, 2708 (2019).

59. R. Quentin, L. Fanuel, M. Kiss, M. Vernet, T. Vékony, K. Janacsek, L. G. Cohen, D. Nemeth, Statistical learning occurs during practice while high-order rule learning during rest period. *npj Science of Learning* **6**, 14 (2021).

60. E. A. McDevitt, J. Zhang, K. J. MacKenzie, J. Fiser, S. C. Mednick, The effect of interference, offline sleep, and wake on spatial statistical learning. *Neurobiol. Learn. Mem.* **193**, 107650 (2022).

61. R. Stickgold, M. P. Walker, Sleep-dependent memory triage: evolving generalization through selective processing. *Nat. Neurosci.* **16**, 139–145 (2013).

62. J. Arato, C. A. Rothkopf, J. Fiser, Learning in the eyes: specific changes in gaze patterns track explicit and implicit visual learning, *bioRxiv* (2020). https://doi.org/10.1101/2020.08.03.234039.

63. J. M. Ellenbogen, P. T. Hu, J. D. Payne, D. Titone, M. P. Walker, Human relational memory requires time and sleep. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 7723–7728 (2007).

64. S. Fischer, S. Drosopoulos, J. Tsen, J. Born, Implicit Learning–Explicit Knowing: A Role for Sleep in Memory System Interaction. *J. Cogn. Neurosci.* **18**, 311–319 (2006).

65. V. Daelli, N. J. van Rijsbergen, A. Treves, How recent experience affects the perception of ambiguous objects. *Brain Res.* **1322**, 81–91 (2010).

66. L. Shams, U. R. Beierholm, Causal inference in perception. *Trends Cogn. Sci.* **14**, 425–432 (2010).

67. X.-X. Wei, A. A. Stocker, A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. *Nat. Neurosci.* **18**, 1509–1517 (2015).

68. I. Djonlagic, A. Rosenfeld, D. Shohamy, C. Myers, M. Gluck, R. Stickgold, Sleep enhances category learning. *Learn. Mem.* **16**, 751–755 (2009).

69. D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, D. J. Kupfer, The Pittsburgh

sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).

70. T. F. Meijman, A. H. de Vries-Griever, G. de Vries, *The Evaluation of the Groningen Sleep Quality Scale* (Heymans Bulletin, Groningen, 1988).

# Supplementary Material

## Supplementary Experiment 1

*Participants*

243 participants (128 female, Age: mean = 30.1, mode = 24, SD = 11.7) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 1. The study was approved by the Psychological Research Ethics Board of the Central European University and all participants provided informed consent.

*Materials*

This experiment used the same materials as Experiment 1. However, the shapes were not grouped into horizontal and vertical pairs but into two orthogonal groups of oblique pairs (Figure 4). For one group of pairs, the second shape was always in the top right grid cell from the first shape, for the other group of pairs the second shape was always in the top-left grid cell. As such pairs allow for fewer unique combinations within a 3x3 grid the scenes for this Experiment were set in a 5x5 grid. The shapes mainly occupied the central 3x3 sub-grid, with one shape per scene being in the outer cells. All shapes appeared in the outer cells an equal number of times over all scenes. In order to assure that the whole grid is visible on the participants' devices, the size was not set to a fixed pixel value. Instead, the size of one grid cell was set to 1/7 of the pixel height of the participant’s screen. Therefore, the whole grid filled 5/7 of the screen height.

*Procedure*

The procedure was identical to Experiment 1.

*Results*

The exclusion criteria were identical to those used in Experiment 1. This led to 15 exclusions for failed attention checks and 4 exclusions for response bias. This left us with 224 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in Experiment 1. This led to 12 participants being categorized as explicit and 212 as implicit. Bayes Factors from Bayesian t-tests for implicit participants reported for experiments 1, 2, and 3 used an r-scale parameter of 0.5 instead of the default √2/2, reflecting that Experiment 1 found small effect sizes for this group.

Overall, the data shows the same pattern as in Experiment 1. The results for the implicit participants (n=212) closely follow the results of Experiment 1. They perform above chance for pairs of the first training phase (M=53.9, SE=0.8, d=0.34, t(211)=4.99 p<0.001, BF=10258) and for pairs of a novel structure (M=57.1, SE=1.7, d=0.28, t(211)=4.06 p<0.001, BF=245) but not pairs of the same structure (M=50.1, SE=1.8, d=0.03, t(211)=0.46 p=0.999, BF=0.12) in the second training phase. The performance for pairs of the same and novel structure is again significantly different (d=0.18, t(211)=2.67 p=0.049, BF=3.2). The results for explicit participants (n=12) show the same qualitative pattern as in Experiment 1, however without reaching a significant difference from chance: first training phase: M=59.4, SE=4.8, d=0.56, t(11)=1.95 p=0.387, BF=1.2), novel structure: M=52.1, SE=10.4, d=0.06, t(11)=0.20 p=0.999, BF=0.29), same structure: M=60.4, SE=7.8, d=0.39, t(11)=1.33 p=0.084, BF=0.59). This can be explained by the significantly smaller number of participants acquiring explicitness in Experiment 5 as compared to 1 ($X^2$=10.47, df=1, p=0.001, BF=37.0) which results in diminished power for these tests.

## Supplementary Analysis Details

### *Experiment 1*

*Participants*

The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB) and all participants provided informed consent. The hourly compensation was £ 6.3 .The sample size was chosen to achieve 80% power for expected small effect sizes (d=0.2) in paired t-tests (needed sample according to power analysis = 198.15) and to account for exclusions.

Based on pilot data we choose 20 seconds combined response time for both attention checks as the cut-off value for inclusion. 19 participants were rejected for failing this criterion. Response bias was defined as the proportion with which participants used one of the two response options (“1” and “2”) and participants who were 2.5 SD away from the mean were excluded. 3 participants were excluded for failing this criterion. This left us with 229 participants after exclusions.

*Materials*

As this was an online study, participants conducted it on their own computers using google chrome, safari, or opera browser. Only desktop and laptop computers were admissible, and no smartphones or tablets. Stimuli were presented using custom JavaScript code built on the jsPsych library[1]. As participants used different devices (screen size and resolution) the visual angle of the shapes was not the exact same for all participants. Instead, the 3x3 grid extended over 600x600 pixels and was centered in the middle of the screen. The remaining screen outside the grid was empty (white).

*Procedure*

Participants passively observed 160 scenes in the first training phase. For half of the participants, these scenes contained only horizontal pairs (horizontal condition), and for the other half only vertical pairs (vertical condition). Each scene was presented for 2 s with a 1 s interstimulus interval (ISI). After a two-minute passive break, participants passively observed 96 scenes in the second training phase. Participants were not told about the presence of any structure in the scenes and were simply instructed to be attentive so that they could later answer simple questions. After half of each training phase, an attention check appears, asking participants to press the spacebar to continue. Response time for the attention check was recorded to detect inattentive participants. After the second training phase, participants had another two-minute passive break. Following this, pair learning was tested with a two-alternative forced choice task (2AFC). In each trial, participants saw a real shape pair from one of the training phases and a foil pair created by combining shapes from two different pairs of the same training phase. Overall, all real and foil pairs were used the same number of times during the test phase to ensure no learning effects within the test phase. Real and foil pairs were presented after each other in the 3x3 grid for 2 s with a 1 s ISI. The order of real and foil pairs was randomized. Participants were asked to indicate which of the two was more familiar by pressing “1” or “2” on their keyboard. Participants first completed 16 trials using pairs from the second training phase, followed by 24 trials using pairs from the first training phase. Finally, participants answered five open questions about their beliefs about the experiment and their knowledge of pair structure.

*Results*

The data was collapsed over vertical and horizontal conditions for all further analysis, as a 3x2 mixed ANOVA with *test type* (levels: training 1, same structure, novel structure) as within-subject factor and *condition* (levels: horizontal, vertical) as between-subject factor

showed no significant main effect of condition (F(1, 224)=0.776, p=0.379, BF= 0.08, $\eta_p^2$=0.003) and no significant *test type* - *condition* interaction (F(2,448)=0.199, p=0.820, BF= 0.04, $\eta_p^2$=0.001).

Bayes Factors (BF) reported for the ANOVA here and throughout the text are based on Bayesian ANOVAs using the *BayesFactor* R package, realizing Bayesian tests with models analogous to the frequentist counterpart, and employing a JZS prior[2]. BFs reported for t-tests throughout the text are calculated with the same package, following the same logic, and again employing the JZS prior unless specified otherwise. All other settings for the R functions of the *BayesFactor* package where left at the default values. P-values reported throughout the text were subject to experiment-wise correction for multiple comparisons using the Holm-Bonferroni method. All conducted tests are two-sided, unless otherwise specified.

To test for a possible time-of-day effect in learning or generalization we correlated test performance with the hour of the day at which participants completed the experiment. There were no significant correlations for pairs of the same structure (explicit participants: r=-0.03, p=0.882; implicit participants: r=0.01, p=0.893) or pairs of the novel structure (explicit participants: r=0.13, p=0.477; implicit participants: r=-0.05, p=0.458). Additionally, we looked separately at groups of participants completing the experiment early in the day (7-11 am) and late in the day (7-11 pm). For implicit participants, there was no significant difference between participants that participated early (n=23) or late (n=22) as a 2x2 mixed ANOVA with *hour-of-day* and *test type* as factors showed no significant main effect of *hour-of-day* (F(1,43)=0.019, p=0.892, BF=0.26) and no significant *hour-of-day – test type* interaction (F(1,43)=0.095, p=0.759, BF=0.31).

***Experiment 2***

*Participants*

The sample size was chosen to match that of Experiment 1. The study was approved by the Psychological Research Ethics Board of the Central European University and all participants provided informed consent. The hourly compensation was £ 6.3. To ensure that participants have overnight sleep during the experiment as intended several constraints and checks were implemented (see *Sleep During Consolidation Studies* section of the supplementary material).

The same exclusion criteria as Experiment 1 were used. This led to 18 exclusions for failed attention checks and 2 exclusions for response bias. In addition to the exclusion criteria used in Experiment 1, here we also employed exclusion criteria related to sleep quality. 47 participants were excluded because they reported bad sleep quality for the night before the experiment or the night of the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). Additionally, 31 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). This left us with 161 participants after exclusions

***Experiment 3***

*Participants*

The sample size was chosen to match that of Experiment 1. The study was approved by the Psychological Research Ethics Board of the Central European University and all participants provided informed consent. The hourly compensation was £ 6.3. To ensure that participants have overnight sleep during the experiment as intended several constraints and checks were implemented (see *Sleep During Consolidation Studies* section of the supplementary material).

The same exclusion criteria as Experiment 2 were used. This led to 29 exclusions for failed attention checks and 3 exclusions for response bias. 28 participants were excluded because they reported bad sleep quality for the night before the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). 28 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). Additionally, 17 participants in this experiment were excluded as they reported that they slept at all during the day. This left us with 170 participants after exclusions.

***Experiment 4***

*Participants*

The sample size was chosen to match that of Experiment 1. The study was approved by the Psychological Research Ethics Board of the Central European University and all participants provided informed consent. The hourly compensation was £ 6.3. To ensure that participants have overnight sleep during the experiment as intended several constraints and checks were implemented (see *Sleep During Consolidation Studies* section of the supplementary material).

The same exclusion criteria as Experiment 2 were used. This led to 17 exclusions for failed attention checks, and 5 exclusions for response bias. 51 participants were excluded because they reported bad sleep quality for the night before the experiment or the night of the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). Additionally, 33 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). This left us with 169 participants after exclusions.

### *Experiment 5*

*Participants*

The sample size was chosen to approximately match the number of explicit participants in Experiment 1 after exclusions. The study was approved by the Psychological Research Ethics Board of the Central European University and all participants provided informed consent. The hourly compensation was £ 6.3. To ensure that participants have overnight sleep during the experiment as intended several constraints and checks were implemented (see *Sleep During Consolidation Studies* section of the supplementary material).

The exclusion criteria were identical to those used in Experiment 1. This led to 4 exclusions for failing the attention checks, and 0 exclusions for response bias. This left us with 36 participants after exclusions.

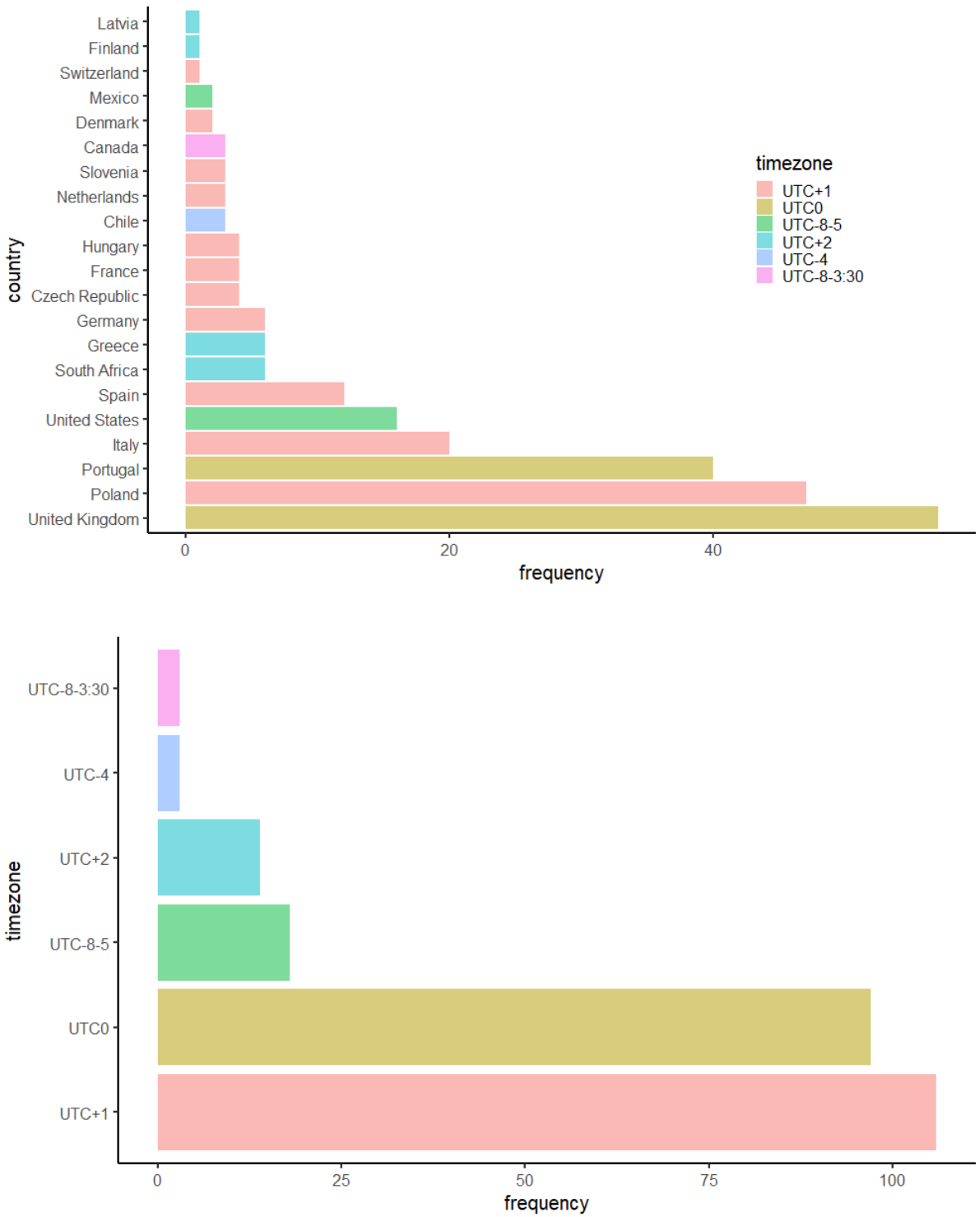
### *Analysis Across Experiments*

To compare directly the differential effect of type of consolidation on implicit structure learning, we entered the data of participants with implicit knowledge from Experiments 1, 2, 3 and 4 into a 4x2 ANOVA, with *consolidation type* (no consolidation, 12-h-sleep, 12-h-awake, and 24-h-sleep consolidation) and *structure type* (same or novel structure) as factors. The obtained results showed the typical pattern of a cross-over interaction with no significant main effects (*consolidation type*: $F(3,623)=0.18$, $p=0.910$, $BF=0.003$, $\eta_p^2=0.0009$; *structure type*: $F(1,623)=1.52$, $p=0.218$, $BF=0.17$, $\eta_p^2=0.002$) but a significant interaction ($F(3,623)=7.43$, $p<0.001$, $BF=1979$, $\eta_p^2=0.03$). Post-hoc tests revealed significant differences between the no-consolidation group (Exp. 1) and the two asleep-consolidation groups (Exp. 3 and 4), where the no-consolidation group showed stronger learning of novel structure pairs (Exp. 1 vs Exp. 2: $p=0.004$, $BF=44.3$; Exp. 1 vs Exp. 4: $p=0.012$; $BF=12.7$), while the asleep-consolidation groups showed stronger learning of same structure pairs (Exp. 1 vs

Exp. 2: p=0.015, BF=8.8; Exp. 1 vs Exp. 4: p=0.011; BF=14.3). No other significant differences were found.

**Sleep During Consolidation Studies**

In order to assure that participants had overnight sleep during Experiments 2 and 4, and that they didn't sleep during the day in Experiment 3, a number of constraints and checks were implemented. First, participants were not taken from the full global prolific pool, but restricted to a number of European countries within the same time zones. Country of residence is one of the attributes of participants that prolific.co verifies. In order to roughly approximate the geographic distribution of participants in Experiment 1 (see Supplementary Figure 1), we chose countries from two time zones. GMT±00:00 (Countries: UK and Portugal) and GMT+01:00 (Countries: Germany, France, Spain, Czech Republic, Denmark, Hungary, Italy, Netherlands, Poland, Slovenia, Switzerland). Second, as we don’t expect Prolific’s residence information to be perfectly predictive of where participants are while they conduct the experiment, participants were asked what the current time at their location is. Third, at the start of the second part of the experiment, participants were asked how much they slept between the first and second parts. This was used to exclude participants from Experiment 3 who slept during the day.




*Supl. Fig. 1. Country of Residence and Time Zone for Participants in Experiment 1.*

**Matched Sample Analysis**

As reported in the main text, explicit participants show higher average learning in the first learning phase, which could be what enables the generalization of the learned structure. To test this idea, we conducted a matched sample analysis[3]. The general idea of this analysis is to create a sub-sample of the implicit participants that perform like the explicit participants for the pre-training trials.

For all experiments, In a first step, we ran six applicable matching algorithms implemented in the MatchIt R package[4]. The six so-created matched implicit samples were then compared to the original explicit sample according to four metrics: standardized mean difference, variance ratio, mean of the empirical cumulative density function, and maximum of the empirical cumulative density function. All values for all experiments can be seen in Supplementary Tables 1-4. “Unbalanced” denotes the values for the full, non-matched implicit sample. All values for the used matching methods are evaluated as an improvement from those values. Std. Mean Diff. describes how far the mean of the matched sample is from the comparison sample (explicit participants); values closer to zero are better. The variance ratio is the ratio of the variances of the matched and the comparison sample; the best possible value is 1. The eCDF (empirical cumulative density function) contains more information than the mean and variance ratio as they capture the whole distribution of values. Two commonly used simple metrics based on the eCDF are the mean and maximum difference of the eCDFs of the matched and comparison group. Generally, values closer to zero are better. The best-fitting matching algorithm was not exactly aligned for all experiments. For consistency reasons, we chose the overall best-fitting method for all experiments: nearest neighbour matching with replacement.

***Experiment 1 Results***

We found that the matched sample showed a similar pattern for learning in the second training phase as the original full sample (see Figure 3). This is captured by a 2x2 ANOVA using the novel and same structure pairs for the original explicit and the matched implicit data showing a significant interaction ($F(1,87)=8.53$, $p=0.004$, $BF=10.7$, $\eta_p^2=0.09$) and post-hoc comparisons show a significant difference between novel and same structure trials for the synthetic implicit data ($p=0.012$; $BF=3.6$). This analysis suggests that the difference between the two groups is not merely based on different strengths of learning in the first training phase.

***Experiment 2 Results***

As in Experiment 1, the matched sample showed the same type of pattern as the full sample (see Figure 3). As a critical analysis, we can see that for the matched implicit sample there is a significant difference between learning pairs of the same and of the novel structure ($d=0.98$, $t(20)=4.51$, $p<0.001$, $BF=127$), suggesting generalization of the structure.

***Experiment 3 Results***

As previously, the matched sample showed a similar type of pattern as the full sample (see Figure 3). Critically, we can see that for the matched implicit sample there is no significant difference between learning pairs of the same and of the novel structure ($d=0.29$, $t(19)=-1.29$, $p=0.214$, $BF=0.59$), suggesting no generalization of the structure.

***Experiment 4 Results***

As in Experiment 1, the matched sample showed the same type of pattern as the full sample descriptively (see Figure 3). However, the critical analysis of difference between learning pairs of the same and of the novel structure for the matched implicit sample failed to reach significance ($M_{diff}=8.66$, $d=0.22$, $t(22)=1.05$, $p=0.304$, $BF=0.46$).

**Supplementary Table 1**

*Experiment 1 - Overview of Balance Metrics for the Used Matching Algorithms*

| Matching Method | Std. M. Diff. | Var. Ratio | eCDF mean | eCDF max |
|---|---|---|---|---|
| Unbalanced | 0.486 | 2.92 | 0.169 | 0.358 |
| **NN with replacement** | **-0.002** | **0.95** | **0.003** | **0.059** |
| NN without replacement | 0.078 | 1.31 | 0.026 | 0.206 |
| Optimal pair matching | 0.078 | 1.31 | 0.026 | 0.206 |
| Optimal full matching | -0.018 | 1.05 | 0.012 | 0.059 |
| Coarsened exact matching | -0.021 | 1.12 | 0.021 | 0.118 |
| Subclassification | -0.049 | 1.19 | 0.020 | 0.059 |

**Supplementary Table 2**

*Experiment 2 - Overview of Balance Metrics for the Used Matching Algorithms*

| Matching Method | Std. M. Diff. | Var. Ratio | eCDF mean | eCDF max |
|---|---|---|---|---|
| Unbalanced | 1.125 | 2.909 | 0.273 | 0.548 |
| **NN with replacement** | **0.432** | **2.424** | **0.105** | **0.429** |
| NN without replacement | 0.103 | 1.089 | 0.030 | 0.238 |
| Optimal pair matching | 0.432 | 2.424 | 0.105 | 0.429 |
| Optimal full matching | 0.112 | 1.079 | 0.033 | 0.238 |
| Coarsened exact matching | 0.142 | 1.130 | 0.037 | 0.294 |
| Subclassification | 0.248 | 1.175 | 0.060 | 0.238 |

**Supplementary Table 3**

*Experiment 3 - Overview of Balance Metrics for the Used Matching Algorithms*

| Matching Method | Std. M. Diff. | Var. Ratio | eCDF mean | eCDF max |
|---|---|---|---|---|
| Unbalanced | 1.067 | 3.284 | 0.290 | 0.49 |
| **NN with replacement** | **0.293** | **2.260** | **0.079** | **0.30** |
| NN without replacement | 0.043 | 0.950 | 0.018 | 0.25 |
| Optimal pair matching | 0.293 | 2.260 | 0.079 | 0.30 |
| Optimal full matching | 0.068 | 1.073 | 0.023 | 0.25 |
| Coarsened exact matching | 0.043 | 1.161 | 0.023 | 0.25 |
| Subclassification | 0.089 | 1.106 | 0.039 | 0.25 |

**Supplementary Table 4**

*Experiment 4 - Overview of Balance Metrics for the Used Matching Algorithms*

| Matching Method | Std. M. Diff. | Var. Ratio | eCDF mean | eCDF max |
|---|---|---|---|---|
| Unbalanced | 0.819 | 5.352 | 0.222 | 0.494 |
| **NN with replacement** | **0.450** | **3.006** | **0.106** | **0.391** |
| NN without replacement | 0.164 | 1.327 | 0.043 | 0.304 |
| Optimal pair matching | 0.450 | 3.006 | 0.106 | 0.391 |
| Optimal full matching | 0.151 | 1.479 | 0.047 | 0.304 |
| Coarsened exact matching | 0.026 | 1.099 | 0.016 | 0.064 |
| Subclassification | 0.240 | 2.404 | 0.071 | 0.304 |

**Supplementary References**


1. Leeuw, J. R. de. jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* **47**, 1–12 (2015).
2. Rouder, J. N., Morey, R. D., Speckman, P. L. & Province, J. M. Default Bayes factors for ANOVA designs. *J. Math. Psychol.* **56**, 356–374 (2012).
3. Ho, D. E., Imai, K., King, G. & Stuart, E. A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Polit. Anal.* **15**, 199–236 (2007).
4. Ho, D. E., Imai, K., King, G. & Stuart, E. A. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Softw.* **42**, (2011).