INTRODUCTION

Recently acquired information is often susceptible to distortion and loss. A subset of novel memories must thus undergo an active process of consolidation (Squire et al., 1984). This process, often referred to as ‘systems consolidation’, occurs mainly during post-learning sleep (Klinzing et al., 2019). Successful consolidation reduces the probability of information to be forgotten. It also induces qualitative changes to the information, often including improved generalisation. For visual learning, this may encompass a higher invariance to specific properties of a stimulus, such as its location in the visual field.

Visual learning is commonly investigated using the classical texture discrimination task (TDT). Improvements on this task tend to be confined to the trained eye or region of the visual field. This suggests that visual learning is supported by synaptic changes in the early stages of processing, where information is retinotopically organised (Karni & Sagi, 1991). This notion is supported by blood-oxygen-level-dependent responses to the TDT, which are higher in the visual cortex when stimuli are presented to the trained compared to the untrained eye (Schwartz et al., 2002). A series of studies showed benefits of sleep on TDT improvements (Deliens et al., 2014; Gais et al., 2000; Karni et al., 1994; Stickgold et al., 2000).
et al., 2000; Tamaki et al., 2020), including the ability to generalise to the untrained eye (Deliens et al., 2014). Due to the retinotopic organisation of early visual areas, local mechanisms are unlikely to account for this effect, suggesting a role of top-down feedback during consolidation. Coordinated memory reactivation between visual cortex and hippocampus (Ji & Wilson, 2007), as well as thalamus (Durkin et al., 2017), has been shown in sleeping animals, potentially reflecting such a consolidation mechanism.

As a major caveat to this idea, sleep benefits on visual perceptual learning have almost exclusively been demonstrated on the TDT. Importantly, difficulty in the TDT is adjusted by changing the time between the stimulus and a subsequent mask, not by changing properties of the stimulus itself. Improvements on the task may therefore result from better temporal allocation of attentional resources. Consistent with this idea, training subjects on temporal aspects of the task beforehand has been shown to minimise further improvements (Wang et al., 2013) and to improve generalisation (Xiong et al., 2016). This suggests that learning and consolidation in the TDT may involve higher levels of visual processing than is commonly assumed.

To specifically investigate perceptual learning and its generalisation at the lowest levels of the visual system, we employed a task that trains the ability to assess depth by using disparities in the images received by both eyes. First signs for processing of binocular disparity are found in the primary visual cortex (DeAngelis, 2000). We define perceptual learning as retinotopically local performance improvements following repeated exposure to this low-level task. We define generalisation as performance improvements at untrained retinotopic locations. Note that this view of generalisation is restricted in that it does not account for the possibility that learning may generalise to new retinotopic locations, while remaining specific to the trained location in external space. Furthermore, we assessed generalisation in terms of immediate post-consolidation improvements in performance at the untrained location. However, generalisation may also be reflected in faster performance gains at that location. As is the case for the seminal studies described above, our protocol does not allow distinguishing between these different forms of generalisation.

In a previous study using the same task, we did not find evidence of sleep-associated retinotopically local improvements or their generalisation across the horizontal meridian of the visual field (Klinzing et al., 2020). However, our previous results might have been confounded by differences in baseline performance between the lower and upper hemifield (Skrandies, 1987), such that worse performance in the upper visual field could have masked generalisation. To exclude such confounding, we tested in the present study generalisation across the vertical instead of the horizontal meridian, i.e. from the left to the right visual field or vice versa. Furthermore, we addressed potential carry-over effects between wake and sleep conditions by testing this factor across groups instead of within subjects. Extending the results from our previous study, we present evidence against a generalisation of binocular disparity learning across the vertical meridian. While this does not exclude the existence of local generalisation to close-by receptive fields, our data suggest that early perceptual learning devoid of temporal cues is neither subject to long-range generalisation by top-down feedback, nor does it depend on sleep.

2 | METHODS

2.1 | Participants

A total of 32 participants (17 female, 15 male; mean [± SD, range] age 23.66 [3.25, 18–30] years) were trained on a binocular disparity discrimination task used previously (Klinzing et al., 2020). Nine additional subjects started the experiment but were excluded because they were not able to see the depth effect and thus performed at around chance level (n = 7), slept during the retention interval despite being assigned to the wake group (n = 1), or as a result of technical difficulties (n = 1). Subjects were required to score >1.0 in a decimal visual acuity test, corresponding to a Snellen acuity of 20/20. Exclusion criteria were health problems, ongoing medication, medical interventions, night or shift work, examination periods, stress-intense occupations during the 3 weeks prior to the experiment, or a history of psychiatric, neurological, or sleep disorders. On experimental days, extensive physical exercise, daytime naps, as well as the consumption of alcohol, caffeine, or illegal drugs were prohibited. The study was conducted in accordance with the Declaration of Helsinki and approved by the local Ethics Committee of the Medical Faculty of the Universität Tübingen. All subjects gave their written informed consent.

2.2 | Binocular disparity discrimination task

Participants performed a two-choice disparity discrimination task in the lower visual field (Figure 1a). They discriminated whether a central disc (3° diameter) was protruding (“near”) or receding (“far”) relative to a ring (1° wide) surrounding the disk. Trials were performed at different difficulty levels, varied by the disc’s signal strength (see below). Between trials, only a fixation cross was shown at the centre of the screen. After initiating a trial with a button press, the stimulus (disc and ring) was shown for 1.5 s with a horizontal (left or right depending on condition) and vertical offset (lower visual field) to the fixation cross of 3°. The stimulus and fixation cross then disappeared and were replaced for a maximum of 2 s by two-choice targets (small versions of a near and a far stimulus at full signal strength; 2.2° disc diameter; 0.8° ring width) above and below the previous location of the fixation cross. Participants were asked to pick the stimulus they had just seen using an up or down button press (match-to-sample). Eye fixation was assessed using an infrared eye tracker during the interval between stimulus onset and choice target onset (Eyelink 1000, SR Research) at a sampling rate of 500 Hz. Stimuli were shown on a back-projection screen using two projectors (Projection Design F21 DLP, 60 Hz,
1,920 × 1,080 pixels, 70.5 cm image width, 225 cd/m² mean luminance, linearised, 80 cm viewing distance) and passive linear polarising filters with a relative tilt of 90°. Participants wore passive linear polarising filter glasses, allowing us to present slightly different images to each eye. The disparity between the two images resulted in the perception of depth, which we used to adjust the relative distance of the central disc to the outer ring.

The disc and ring were assembled from circular dynamic random dot stereograms, consisting of equal numbers of black and white dots. The outer ring was always shown at 0° disparity. For each video frame (shown for 1/60 s), all dots belonging to the central disc had identical disparities. Target disparities (±0.1° for “near” and “far”) were well above the detection threshold and thus easily detectable in the absence of noise. Task difficulty was manipulated by varying the proportion of video frames showing dots at the target disparity (% signal strength). For the remaining video frames, disparity values were randomly drawn from a uniform distribution of 11 values in 0.05° increments from −0.25° to 0.25°. For 0% signal trials, disparity values for all video frames were drawn from the noise distribution and random feedback was given. We ran the task using custom scripts (cf. Seillier et al., 2017) using MATLAB 2014a (MathWorks) and Psychophysics Toolbox 3 (Kleiner et al., 2007; Peli, 1997).

Performance was modelled using cumulative Gaussian functions. The resulting psychometric curves describe the ratio of “far” answers (represented on one axis) for each of the sampled signal strengths (represented on the other axis). At one extreme, trials with a full-strength “near” signal usually lead to almost no “far” answers. At the other extreme, trials with a full-strength “far” signal lead to almost exclusive “far” answers. For these and all signal values in between, the ratio of “far” answers can be adequately captured by a sigmoid curve. From this psychometric function, three parameters were derived: the Slope of the function is the main metric for perceptual performance. Assessed at the signal strength leading to 50% correct responses, it reflects how quickly performance improves with higher (and declines with lower) signal strength. A related measure is the psychometric Threshold, which is defined as the signal strength required for 85% performance (here we use the average of the unsigned values for near and far trials). The third parameter we analysed is the observer’s Bias. On trials where no signal is given, an unbiased observer would make the same amount of “far” and “near” decisions. For real observers, this equilibrium is usually located at a signal strength that deviates from 0. This signal strength is considered the observer’s Bias.

2.3 | Experimental timeline

All subjects participated in an Acquisition and a Testing session, which were scheduled 12 hr apart and consisted of three phases
each (Figure 1c). During Acquisition, we introduced participants to the task in a Warmup phase, during which stimuli were displayed with high signal strength (120 trials; signal at 85% or 95%). During a Baseline phase, we linearly sampled the range of available signal strengths to assess the participant's performance (200 trials; signal at 0%, 20%, 40%, 60%, and 80%). Finally, in the Training phase, we presented stimuli with lower signal strengths, depending on their performance during Baseline. In addition to these rather difficult trials, we added one very high but rarely sampled signal strength to assess performance during Baseline. In addition to these rather difficult trials, we added one very high but rarely sampled signal strength to assess the subject's response rate (Wichmann & Hill, 2001; 800 trials with lower signal strengths being sampled more densely, e.g. 5%, 10%, 20%, 40%, and 90%).

During Testing after the 12-hr retention period, we briefly re-familiarised subjects with the task (Refresher, 16 trials; signal at 85% and 95%). Afterward, Retrieval and Generalisation were assessed with signal strengths identical to Training (280 trials each). During Generalisation, the stimulus was presented again in the lower visual field but on the other side of the vertical meridian (i.e. moved from the left to the right side or vice versa, Figure 1b). Trials were aborted in case of a fixation break during Warmup (after 40 initial trials that allowed free exploration), as well as during Baseline and the Refresher phase. When trials were aborted, a tone notified the participants about the fixation break. During all other phases of the experiment, trials with fixation breaks were completed regularly but data from these trials were discarded from analysis. To increase motivation, written feedback on task performance and fixation accuracy was given every 40 trials. Trial-by-trial auditory feedback signalling correct versus incorrect responses was only provided during Acquisition and disabled during Testing to minimise further learning.

Participants in the sleep group performed Acquisition in the evening, slept during the retention interval (mean [SD] 7.48 [0.72] hr), and went through Testing in the morning. Participants in the wake group started Acquisition in the morning, stayed awake, and completed Testing in the evening. We verified that subjects in the wake group did not nap during retention using activity trackers with acceleration and light sensors (Actiwatch, Philips Respironics). The same devices were used to assess sleep duration in the sleep group. Any sleep in the wake group or sleep for <6 hr in the sleep condition led to the exclusion of the subject (n = 1). Subjects showed an increase in sleepiness across a wake retention interval (Holm-corrected p = .011, d = 3.469) and no change across a sleep retention interval (p = 1.0; interaction Acquisition/Testing x Sleep, p = .011, $\eta^2 = 0.079$).

### 2.4 Data analysis

We employed Bayesian inference (psignifit toolbox for Matlab, https://github.com/wichmann-lab/psignifit) for estimating psychometric functions for Baseline, Training, Retrieval, and Generalisation by fitting a beta-binomial model to the correct responses at each signal strength (Schütt et al., 2016). Based on the fitted cumulative Gaussian functions (see “Binocular disparity discrimination task”), we analysed the parameters Slope, Threshold, and Bias. On average, our models explained 96.55% of the variance in the data (minimum: 84.70%). Lapse rates were incorporated in the model and were generally very low (mean [SD] across subjects and experimental phases $2.17 \pm 2.44\%$).

Only trials in which participants correctly maintained fixation were analysed. Outlier rejection was performed following an interquartile range outlier rejection rule with a multiplier of 2.2 (lower threshold: $Q1 - 2.2 \times [Q3 - Q1]$; upper threshold: $Q3 + 2.2 \times [Q3 - Q1]$). Statistics were performed using JASP 0.13 (https://jasp-stats.org) and by calculating mixed repeated measures analysis of variance (ANOVAs) incorporating the within-subject factor Phase (Baseline/Training/Retrieval/Generalisation) and the between-subject factor Wake/Sleep. Performance changes throughout the experiment were assessed based on the factor Phase, while an effect of sleep on consolidation would be reflected by an interaction of the factors Phase and Wake/Sleep. Direct comparisons between phases or groups were performed using paired or unpaired samples t tests, respectively. In an additional analysis, we split the Training phase into three blocks of 260 trials each (Training 1–3, the first 20 trials were discarded). We confirmed the results of the classical analysis by conducting additional Bayesian statistics. As the first night of a previous experiment (Klinzing et al., 2020) was identical to the present study, we aggregated data across studies for a second Bayesian analysis.

Post hoc tests were corrected using Holm’s method. In case the sphericity assumption was violated, results were Greenhouse–Geisser corrected (original degrees of freedom are then reported together with the correction factor $\epsilon$). The non-parametric Friedman test was used in one case (correct fixations across experimental phases) in which Levene’s test showed unequal variances. Effect sizes are provided for significant tests ($\eta^2$ for ANOVA, Cohen’s $d$ for post hoc tests). Data were visualised using Python 3.7 and Seaborn 0.10.1. Error bars are shown as bootstrapped 95% confidence intervals and, in case of skewed bootstrapped distributions, may be asymmetric. Error bars for Bayesian analyses reflect 95% credible intervals.

### 3 RESULTS

#### 3.1 Performance in the binocular disparity task improves locally

Task performance, as measured by the slope of the psychometric function, changed over the course of the experiment across both groups (mixed repeated measures ANOVA, main effect Phase, $F_{3,90} = 9.818$, $p < .001$, $\eta^2 = 0.092$). Performance increased from Baseline to Retrieval, which was reflected by a progressive steepening of the slope (Figure 2, top). Performance data showed a trend towards improvement from Baseline to Training (post hoc paired-samples t test on data collapsed across groups, Baseline versus Training, $t_{31} = 2.034$, $p = .152$, Holm-corrected), a significant increase...
over the retention interval (Training versus Retrieval, $t_{31} = 3.587, p = .006, d = 0.634$), and a significant reduction when generalisation of learning was tested in the opposite quadrant of the lower visual field (Retrieval versus Generalisation, $t_{31} = −3.372, p = .008, d = −0.596$). As a result, performance during Generalisation was not significantly different from Baseline levels, when participants were still naïve to the task (Generalization versus Baseline, $t_{31} = 1.583, p = .268$). In a separate analysis, we analysed performance changes during Training by splitting that phase into three blocks (Figure 2, bottom). Interestingly, participants on average reached high performance levels early during Training and showed no further improvements after the first of the three Training blocks (mixed repeated measures ANOVA, main effect Training block, $F_{2,58} = 0.573, p = .567$). Above findings were corroborated by results based on the psychometric threshold (mixed repeated measures ANOVA, main effect Phase, $F_{2,84} = 8.589, p < .001, \eta^2 = 0.091, \epsilon = 0.749$; performance improvements result in lower thresholds). Mirroring the slope data, threshold values significantly decreased over the retention interval (post hoc paired-samples t test on data collapsed across groups, Baseline versus Retrieval, $t_{30} = −4.969, p < .001, d = −0.907$; Training versus Retrieval, $t_{30} = −3.524, p = .006, d = −0.643$), and increased again when the task was performed in the horizontally opposing visual-field quadrant (Retrieval versus Generalization, $t_{30} = 4.055, p = .002, d = 0.740$; all other $p \geq .332$; all tests Holm corrected).

We further analysed bias, an observer’s inherent affinity towards one or the other response. As expected, participants’ bias changed over the course of the experiment (mixed repeated measures ANOVA, main effect Phase, $F_{3,90} = 4.717, p = .009, \eta^2 = 0.074, \epsilon = 0.757$), with the highest values observed during Baseline, when subjects were still unaccustomed to the task, and subsequent decreases for Training and Retrieval (post hoc paired-samples t test on data collapsed across groups, Baseline versus Training, $t_{31} = −3.266, p = .016, d = −0.577$; Training versus Retrieval, $t_{31} = 1.286, p = .416, d = 0.227$). Bias increased slightly during Generalisation, but this change was significant only when compared to Training, which had the lowest bias values (Training versus Generalisation, $t_{31} = 3.144, p = .018, d = 0.556$; all other $p \geq .29$).

Trials in which participants broke fixation were excluded from analysis. However, if the proportion of such trials differed between groups or phases, this could still affect our results. Analysis of fixation data showed that rates of correct fixation were generally very high (mean [SD] 91.34 [6.95]%). We observed a trend for higher values during Retrieval than during Training and Generalisation (non-parametric Friedman test, main effect Phase, $\chi^2 = 5.213, p = .074$). Additional non-parametric between-group comparisons did not indicate significant differences during any of the phases (all $p > .57$). Our results are thus unlikely to be substantially affected by differences in fixation.

![FIGURE 2](image-url) Sleep does not aid the transfer of improvements in binocular disparity to the other visual field. We compared the slope of the psychometric functions fitted to each participant’s performance between experimental phases and across groups (higher values denote an improvement in performance; for other parameters, see main text). Top: Large light unfilled and dark filled circles show grand averages (±95% confidence intervals) for the wake and sleep group, respectively. Small circles show single-subject data. Bottom: Training progress, illustrated by splitting Training trials into three consecutive blocks of 260 trials each. Performance improved only descriptively from Baseline to Training (post hoc test, $p = .152$, Holm corrected), improved significantly between Training and Retrieval ($p = .006$) and deteriorated when Generalisation was tested in the other visual field ($p = .008$; ANOVA main effect Phase $p < .001$). Performance did not differ between wake and sleep groups (ANOVA main effect Wake/Sleep, $p = .688$; interaction Wake/Sleep × Phase, $p = .505$) or across Training blocks ($p = .567$). **$p < .01$; ***$p < .001$.
3.2 | No signs for improved generalisation after sleep

So far, we have demonstrated that across groups, performance improved with task exposure and over the retention period but dropped to baseline levels when the stimulus was presented on the other side of the lower visual field. By now focussing on differences between groups, we analysed whether sleep affected performance, including the participants’ ability to generalise their improvements to the other visual hemifield. In contrast to our previous study (Klinzing et al., 2020), we tested transfer across the vertical instead of the horizontal meridian. This is a decisive difference because anisotropies between the lower and upper visual field may have confounded our earlier results. Consistent with our previous findings, neither performance nor bias evolved differently across the experiment in the wake and sleep groups (mixed repeated measures ANOVA, main effect Wake/Sleep: Slope, $F_{1,30} = 0.165, p = .688$; Threshold, $F_{1,28} = 0.983, p = .330$; Bias, $F_{1,30} < 0.001, p = .996$; interaction Wake/Sleep x Phase: Slope, $F_{3,90} = 0.786, p = .505$; Threshold, $F_{3,84} = 1.644, p = .198, \eta = .0749$; Bias, $F_{3,90} = 0.510, p = .626, \eta = .0757$). Similarly, a planned direct comparison between the sleep and wake groups during Generalisation showed no significant differences in performance or bias (all uncorrected $p \geq .542$).

Exploratory direct group comparisons of Slope, Threshold, and Bias during Retrieval and Generalisation, as well as their change from Baseline or Training to Retrieval or Generalisation showed no significant differences (18 tests; all uncorrected $p > .05$).

To substantiate the notion that sleep does not influence training and generalisation success, we performed Bayesian ANOVAs on all parameters. As Bayesian inference is based on model comparison, this allowed us to quantify the empirical evidence in favour of the null hypothesis, something that is conceptually impossible using classical statistics. Analyses were conducted using JASP’s default priors ($\alpha$ scale = .5 for fixed effects and 1.0 for random effects).

Given the highly significant main effects of Phase, we included this and its interaction with Phase. These analyses provided moderate priors ($\alpha$ for fixed effects and 1.0 for random effects).

While performance increased significantly over the retention period, this increase did not differ between sleep and wakefulness, and improvements remained local in both experimental groups. Thus, disparity discrimination learning does not seem to be amenable to transfer either via sleep-dependent extraction of higher-level regularities (Zhang et al., 2010), improved temporal allocation of attention (Wang et al., 2013), or other sleep-dependent top-down processes. Results may differ in paradigms in which the temporal structure within trials supports task performance or in which the task involves stimuli that are either static or change on a time scale comparable to natural scenes. It is also worth noting that we defined generalisation in terms of performance improvements that manifest immediately after switching the stimulus to a new location. Alternatively, generalisation may also be seen in terms of enhanced learning potential at the new location (e.g. by steepening the learning curve). The present paradigm is not able to detect this form of generalisation if it does not affect overall performance within the 280 trials of that phase of the task.

We found that behavioural performance improved rapidly and plateaued early on during task execution. Such initial fast learning has also been observed in the original TDT (Karni & Sagi, 1993). Later studies repeatedly showed performance limits or even deterioration of performance, especially during prolonged training (Censor & Sagi, 2009; Ofen et al., 2007). In our data, the significant effect of sleep on the generalisation of improvements in coarse disparity discrimination. This result has now proved consistent across two independent sets of participants (total $n = 49$).

All performance data and statistical analyses have been uploaded to an Open Science Framework repository and are publicly available: https://osf.io/zzkwu/ (doi: 10.17605/OSF.IO/2QKUW).

4 | DISCUSSION

The data reported in the present study confirm our previous findings based on the same stimuli and task (Klinzing et al., 2020). They show substantial improvements in coarse binocular disparity discrimination following a single training session. Our present results also corroborate that these learning effects are retinotopically local, i.e. they do not transfer between different visual hemifields. In our previous study, we investigated transfer from the lower to the upper hemifield (Klinzing et al., 2020). The lack of generalisation observed there could at least partly have been the result of anatomical and functional anisotropies between the upper and lower visual hemifields (Herde et al., 2020; Skrandies, 1987). In the present study, we addressed this potential confound by assessing transfer between horizontal hemifields and observed qualitatively and quantitatively similar effects. We again found that sleep during the retention interval did not affect subsequent performance, neither within the trained visual-field quadrant nor at the transfer location. These results suggest that perceptual learning of coarse disparity discrimination depends on retinotopically specific mechanisms that do not benefit from memory consolidation during subsequent sleep.
increase in performance from baseline to retrieval can to a large extent be attributed to improvements that occurred over the retention interval. Such performance gains could either reflect memory consolidation or a release from training-induced adaptation (Censor et al., 2006; Mednick et al., 2005). Specifically, it has been argued that over-exposure to the stimulus during training may reduce performance within and between sessions on the same day. Sleep may then be required to release this (over-)adapted network state. Based on our manipulation, we cannot assess the relative contributions of consolidation versus release of adaptation. However, in the absence of sleep effects, the most parsimonious explanation for our data is that the passage of time alone is sufficient for adaptation to be released and/or perceptual memories to be consolidated following binocular disparity discrimination learning. The lack of an interaction between offline changes in performance and sleep may appear surprising given a long history of perceptual learning studies that have documented beneficial effects of sleep (Deliens et al., 2014; Gais et al., 2000; Karni et al., 1994; Stickgold et al., 2000; Tamaki et al., 2020). However, practically all demonstrations of sleep effects were obtained with a single experimental paradigm, the TDT (Karni & Sagi, 1991). The nature of this particular task differs substantially from our protocol, as well as from many other tasks classically used to investigate perceptual learning (Aberg et al., 2009; Le Dantec & Seitz, 2012). Most importantly, difficulty in the classical TDT is manipulated, and learning is induced, by reducing the delay between target and mask stimuli. Recent evidence suggests that learning on the TDT largely relates to improved temporal learning (Wang et al., 2013) and that appropriate training procedures can induce transfer of texture discrimination learning. This could explain why a growing number of studies report substantial amounts of ocular and retinotopic transfer (Wang et al., 2012; Xiong et al., 2016; Zhang et al., 2010), in contrast to findings based on the classical paradigm (Karni & Sagi, 1991). A promising avenue for future research would be to use more strictly perceptual low-level tasks, such as the present one, and to manipulate individual stimulus (e.g. static versus dynamic) or task dimensions (e.g. differential rewards) to probe the boundary conditions for learning to become transferable or sleep-dependent. An additional explanation for the lack of sleep effects could be the strict control of eye-movements implemented in the present study. Oculomotor learning is known to benefit from sleep (e.g. Albouy et al., 2006), and it has recently been demonstrated that training and re-activating an unrelated oculomotor memory can enhance TDT performance (Klorfeld-Auslender & Censor, 2019). By effectively suppressing oculomotor learning, our protocol may further reduce the number of memory traces that could potentially benefit from sleep.

The main rationale behind the present follow-up study was to account for the potentially confounding effects of numerous visual-field anisotropies (Abrams et al., 2012; Karim & Kojima, 2010), particularly those between the vertical hemifields (Herde et al., 2020; Previc et al., 1995; Rauss et al., 2009; Skrandies, 1987). For example, differences in spatial and temporal resolution between the upper and lower visual field could have masked high-level transfer of learning as well as sleep benefits in our previous study. Some retinotopic specialisations are even directly linked to disparity processing, such as preferences for crossed versus uncrossed disparities in different parts of the visual field (Grabowska, 1983; Previc et al., 1995). The present data suggest that these anisotropies do not affect the consolidation of binocular disparity discrimination learning. While there are also horizontal asymmetries in visual processing (Grabowska, 1983; Okubo & Nicholls, 2008), these are less likely to exert a substantial influence in the context of our stimuli and task (Breitmeyer et al., 1975; Julesz et al., 1976). Thus, we are confident that the present data, in combination with our previous results (Klinzing et al., 2020), provide strong evidence for retinotopically local learning of disparity discrimination that is independent of both sleep and the particular combination of training and testing locations.

The interpretation of our present findings is limited by the study’s sample size. We cannot exclude that more subtle effects of sleep on perceptual learning evaded detection by our analysis in this and our previous study. We therefore encourage further research in this area, including replications using the same experimental paradigm. Furthermore, our experimental design cannot rule out a confounding influence of circadian rhythm. In the sleep group, retrieval and generalisation were tested in the morning while in the wake group they were assessed in the evening. It is possible that in the sleep group, time of the day of the second session masked improved perceptual skills. This will be an important issue for further investigation. While binocular disparity discrimination is well-suited for our particular research question, it is unclear how far the present results can be generalised to other tasks, such as contrast detection, motion detection, or fine form judgements (Seitz, 2017). Future studies should investigate sleep effects on a wider range of perceptual skills. The overall picture emerging from literally thousands of studies is one of a surprising degree of flexibility in low-level sensory processing, even in the adult brain (Rolfs et al., 2018). However, the boundary conditions for this flexibility are much more restricted than for higher forms of learning. Delineating the conditions necessary and sufficient for low-level learning to be re-processed and generalised will be a crucial step towards integrative models of robust perception (Jacobsen et al., 2018).

ACKNOWLEDGEMENTS
This research was supported by grants from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG): SFB 1233 (project no. 276693517) and a Research Fellowship (project no. 430174808).

CONFLICT OF INTEREST
No conflicts of interest declared.

AUTHOR CONTRIBUTION
JGK and KR designed the study, HN provided experimental setup and code. JGK coordinated the study, carried out the experiments,
and analysed the data. JGK and KR carried out the statistical analyses and drafted the manuscript. All authors reviewed the manuscript and gave final approval for publication.

DATA AVAILABILITY STATEMENT
All performance data and statistical analyses have been uploaded to the Open Science Framework and are publicly available (https://osf.io/zqkuw/).

REFERENCES
Aberg, K. T., Tartaglia, E. M., & Herzog, M. H. (2009). Perceptual learning with Chevrons requires a minimal number of trials, transfers to untrained directions, but does not require sleep. Vision Research, 49, 2087–2094.
Abrams, J., Nizam, A., & Carrasco, M. (2012). Isoeccentric locations are not equivalent: The extent of the vertical meridian asymmetry. Vision Research, 52, 70–78.
Albouy, G., Ruby, P., Phillips, C., Luxen, A., Peigneux, P., & Maquet, P. (2006). Implicit oculomotor sequence learning in humans: Time course of offline processing. Brain Research, 1090, 163–171.
Breitmeyer, B., Julesz, B., & Kropff, W. (1975). Dynamic random-dot stereograms reveal up-down anisotropy and left-right isotropy between cortical hemfields. Science, 187, 269–270.
Censor, N., Karni, A., & Sagi, D. (2006). A link between perceptual learning, adaptation and sleep. Vision Research, 46, 4071–4074.
DeAngelis, G. C. (2000). Seeing in three dimensions: The neuropsychology of stereopsis. Trends in Cognitive Sciences, 4, 80–90.
Deliens, G., Schmitz, R., & Peigneux, P. (2014). Intercocular transfer of perceptual skills after sleep. Journal of Vision, 14, 23.
Durkin, J., Suresh, A. K., Colbath, J., Broussard, C., Wu, J., Zochowski, M., & Aton, S. J. (2017). Cortically coordinated NREM thalamocortical oscillations play an essential, instructive role in visual system plasticity. Proceedings of the National Academy of Sciences of the United States of America, 114, 10485–10490.
Gais, S., Pihlal, W., Wagner, U., & Born, J. (2000). Early sleep triggers memory for visual early discrimination skills. Nature Neuroscience, 3, 1335–1339.
Grabowska, A. (1983). Lateral differences in the detection of stereoscopic depth. Neuropsychologia, 21, 249–257.
Herde, L., Uhl, J., & Rauss, K. (2020). Anatomic and functional asymmetries interactively shape human early visual cortex responses. Journal of Vision, 20, 3.
Jacobsen, J.-H., Behrmann, J., Zemel, R., & Bethge, M. (2018). Excessive invariance causes adversarial vulnerability. 7th Int. Conf. Learn. Represent. ICLR 2019, 1–17. arXiv:1811.00401v4.
Ji, D., & Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. Nature Neuroscience, 10, 100–107.
Julesz, B., Breitmeyer, B., & Kropff, W. (1976). Binocular-disparity-dependent upper–Lower hemifield anisotropy and left–Right hemifield isotropy as revealed by dynamic random-dot stereograms. Perception, 5, 129–141.
Karim, A. K. M. R., & Kojima, H. (2010). The what and why of perceptual asymmetries in the visual domain. Advances in Cognitive Psychology, 6, 103–115.
Karim, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. Proceedings of the National Academy of Sciences of the United States of America, 88, 4966–4970.
Karim, A., & Sagi, D. (1993). The time course of learning a visual skill. Nature, 365, 250–252.
Karim, A., Tanne, D., Rubenstein, B., Askenasy, J., & Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. Science, 265, 679–682.
Kleiner, M., Brainard, D., Pelli, D. G., Ingling, A., Murray, R., & Broussard, C. (2007). What’s new in Psychotoolbox-3. Perception, 36, 1.
Klinzing, J. G., Herbik, L., Nienborg, H., & Rauss, K. (2020). Binocular disparity-based learning is retinotopically specific and independent of sleep. Philosophical transactions of the Royal Society of London. Series B. Biological sciences, 375, 20190463.
Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. Nature Neuroscience, 22, 1598–1610.
Klorfeld-Auslender, S., & Censor, N. (2019). Visual-oculomotor interactions facilitate consolidation of perceptual learning. Journal of Vision, 19, 11.
Le Dantec, C. C., & Seitz, A. R. (2012). High resolution, high capacity, spatial specificity in perceptual learning. Frontiers in Psychology, 3, 1–7.
Mednick, S. C., Arman, A. C., & Boynton, G. M. (2005). The time course and specificity of perceptual deterioration. Proceedings of the National Academy of Sciences of the United States of America, 102, 3881–3885.
Ofen, N., Moran, A., & Sagi, D. (2007). Effects of trial repetition in texture discrimination. Vision Research, 47, 1094–1102.
Okubo, M., & Nicholls, M. E. R. (2008). Hemispheric asymmetries for temporal information processing: Transient detection versus sustained monitoring. Brain and Cognition, 66, 168–175.
Peli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437–442.
Previc, F. H., Breitmeyer, B. G., & Weinstein, L. F. (1995). Discriminability of random-dot stereograms in three-dimensional space. International Journal of Neuroscience, 80, 247–253.
Rauss, K. S., Pourtois, G., Vuilleumier, P., & Schwartz, S. (2009). Attentional load modifies early activity in human primary visual cortex. Human Brain Mapping, 30, 1723–1733.
Rols, M., Murray-Smith, N., & Carrasco, M. (2018). Perceptual learning while preparing saccades. Vision Research, 152, 126–138. https://doi.org/10.1016/j.visres.2017.11.009.
Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. Vision Research, 122, 105–123. https://doi.org/10.1016/j.visres.2016.02.002.
Schwartz, S., Maquet, P., & Frith, C. D. (2002). Neural correlates of perceptual learning: A functional MRI study of visual texture discrimination. Proceedings of the National Academy of Sciences of the United States of America, 99, 17137–17142. https://doi.org/10.1073/pnas.242414599.
Seillier, L., Lorenz, C., Kawaguchi, K., Ott, T., Nieder, A., Pourriah, P., & Nienborg, H. (2017). Serotonin Decreases the Gain of Visual Responses in Awake Macaque V1. Journal of Neuroscience, 37, 11390–11405. https://doi.org/10.1523/JNEUROSCI.1339-17.2017.
Seitz, A. R. (2017). Perceptual learning. Current Biology, 27, R631–R636. https://doi.org/10.1016/j.cub.2017.05.053.
Skrandies, W. (1987). The Upper and Lower Visual Field of Man: Electrophysiological and Functional Differences. Progress in Sensory Physiology (pp. 1–93). Springer.
Squire, L. R., Cohen, N. J., & Nadel, L. (1984). The medial temporal region and memory consolidation: a new hypothesis. In: H. Weingartner, & E. S. Parder (eds.) Memory Consolidation: Psychobiology of Cognition (pp. 185–210). Hove, East Sussex, UK: Psychology Press.
Stickgold, R., James, L., & Hobson, J. (2000). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3, 1237–1238. https://doi.org/10.1038/81756

Tamaki, M., Wang, Z., Barnes-Diana, T., Guo, D. A., Berard, A. V., Walsh, E., Watanabe, T., & Sasaki, Y. (2020). Complementary contributions of non-REM and REM sleep to visual learning. *Nature Neuroscience*, 23(9), 1150–1156. https://doi.org/10.1038/s41593-020-0666-y

Wang, R., Cong, L.-J., & Yu, C. (2013). The classical TDT perceptual learning is mostly temporal learning. *Journal of Vision*, 13(5), 9. https://doi.org/10.1167/13.5.9

Wang, R., Zhang, J.-Y., Klein, S. A., Levi, D. M., & Yu, C. (2012). Task relevancy and demand modulate double-training enabled transfer of perceptual learning. *Vision Research*, 61, 33–38. https://doi.org/10.1016/j.visres.2011.07.019

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313. https://doi.org/10.3758/BF03194544

Xiong, Y.-Z., Zhang, J.-Y., & Yu, C. (2016). Bottom-up and top-down influences at untrained conditions determine perceptual learning specificity and transfer. *Elife*, 5, 1–17. https://doi.org/10.7554/eLife.14614

Zhang, J. Y., Zhang, G. L., Xiao, L. Q., Klein, S. A., Levi, D. M., & Yu, C. (2010). Rule-based learning explains visual perceptual learning and its specificity and transfer. *Journal of Neuroscience*, 30, 12323–12328. https://doi.org/10.1523/JNEUROSCI.0704-10.2010

How to cite this article: Klinzing JG, Nienborg H, Rauss K. Sleep does not aid the generalisation of binocular disparity-based learning to the other visual hemifield. *J Sleep Res*. 2021;30:e13335. https://doi.org/10.1111/jsr.13335