

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Robust perceptual learning of faces in the absence of sleep

Zahra Hussain^a, Allison B. Sekuler^{a,b}, Patrick J. Bennett^{a,b,*}^a Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ont., Canada L8S 4K1^b Centre for Vision Research, York University, Toronto, Ont., Canada

ARTICLE INFO

Article history:

Received 1 July 2008

Received in revised form 27 August 2008

Keywords:

Perceptual learning

Consolidation

Sleep

Face identification

Plasticity

ABSTRACT

This study examines the effects of sleep on learning in a face identification task. Five groups of subjects performed a 1-of-10 face identification task in two sessions separated by 3, 12, and 24 h. Session 1 consisted of four blocks of 105 trials each; Session 2 consisted of eight blocks of trials. All groups exhibited significant improvement in response accuracy within each session. Furthermore, between-session learning – defined as the difference in proportion correct between sessions 1 and 2 – was significant for all groups. Between-session learning was greater in groups that slept between sessions, but the effect was small and affected performance only in the first block of trials in Session 2. Overall, we find that sleep's contribution is a small proportion of the total amount learned in face identification, with improvements continuing to accrue in its absence.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Perceptual and motor skills typically improve with practice over a time-course that consists of at least two components: a fast component in which performance improves during a practice session (Fiorentini & Berardi, 1981; Fahle, Edelman, & Poggio, 1995; Fahle, 1994; Karni et al., 1998; Karni & Sagi, 1993; Poggio, Fahle, & Edelman, 1992), and a slow component in which the benefits of practice accumulate across sessions (Brashers-Krug, Shadmehr, & Bizzi, 1996; Fahle, 1994; Karni et al., 1998; Karni & Sagi, 1993; McKee & Westheimer, 1978; Press, Casement, Pascual-Leone, & Robertson, 2005). This second, slow component, sometimes referred to as latent, or offline learning, is affected by sleep (Atienza, Cantero, & Stickgold, 2004; Fenn, Nusbaum, & Margoliash, 2003; Fischer, Hallschmid, Elsner, & Born, 2002; Gottselig et al., 2004; Karni, Tanne, Rubenstein, Askenasy, & Sagi, 1994; Stickgold, James, & Hobson, 2000a; Stickgold, Whidbee, Schirmer, Patel, & Hobson, 2000b; Walker et al., 2003). Results from Siegel (2001) and Walker and Stickgold (2004), for example, suggest that the REM and SWS stages of sleep are important for consolidating newly-acquired information into long-term memory (but see Song, Howard, & Howard, 2007; Vertes & Eastman, 2000; Vertes, 2004). Sleep is thought to be important for establishing the stimulus-specific benefits of perceptual learning (Karni & Bertini, 1997; Karni et al., 1998; Karni & Sagi, 1993), rather than general benefits that pre-

sumably can transfer across experimental conditions (but see Fenn et al., 2003). Sleep's distinct contribution to latent learning is supported by a study of motor learning, which found that the substantial improvement in performance that is observed across days of testing – i.e., after a night's sleep – is uncorrelated with within-session effects (Walker et al., 2003). However, sleep is not necessary for latent learning, which can occur in the absence of sleep if the gap between testing sessions exceeds 4–6 h (Fischer et al., 2002; Gottselig et al., 2004; Robertson, Pascual-Leone, & Press, 2004; Roth, Kishon-Rabin, Hildesheimer, & Karni, 2005).

In the visual domain, most demonstrations of sleep-related benefits have used a particular kind of texture discrimination task. For example, Karni and Sagi (1991) measured thresholds for discriminating peripherally-viewed horizontal and vertical textured targets, which were embedded in an array of short line segments, by varying the SOA between the offset of the target array and the onset of a mask. Using this task, Karni and Sagi (1993) found that sensitivity in a texture discrimination task increased only after at least 8 h elapsed after the initial training session. Subsequently, it was shown that the effects of practice were abolished by the interruption of sleep, with both REM and SWS contributing to the overall amount learned (Karni et al., 1994; Stickgold et al., 2000a; Stickgold et al., 2000b). Finally, a nap taken in-between successive practice sessions reverses the deterioration that otherwise occurs with repeated testing during the same day (Mednick, Nakayama, & Stickgold, 2003; Mednick et al., 2002; Mednick, Arman, & Boynton, 2005). Using a different task, Schoups, Vogels, and Orban (1995) reported that practice lowered orientation discrimination thresholds measured with circular noise fields rotated about the oblique axis, but only when subjects slept between

* Corresponding author. Address: Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ont., Canada L8S 4K1.

E-mail addresses: hussaiz@mcmaster.ca (Z. Hussain), sekuler@mcmaster.ca (A.B. Sekuler), bennett@mcmaster.ca (P.J. Bennett).

sessions. Learning in both the texture discrimination task and the orientation discrimination task was specific to the trained location in the visual field and, in the orientation discrimination task, to the orientations used during training. For both tasks, within-session gains were only observed early within the first session, after which performance saturated. Further improvements in performance only occurred across sessions, i.e., the improvements were latent.

The role of sleep in consolidation is yet to be demonstrated for the remaining variety of visual tasks amenable to learning. We have previously shown that performance on a 1-of-10 forced choice face-identification task improves substantially across days (Gold, Bennett, & Sekuler, 1999b), and that the improvements are specific to the trained face exemplars and orientations (Hussain, Bennett, & Sekuler, 2005, 2006). Here we assess the contribution and sleep-dependence of latent learning to the overall improvement found with this task, and examine whether the amount of latent learning is associated with within-session learning.

2. Methods

2.1. Subjects

One hundred and three McMaster University undergraduate students participated in this experiment. All subjects had normal or corrected-to-normal Snellen visual acuity. The mean age and years of education were, respectively, 20.9 (SD = 3.31) and 16.9 (SD = 2.30). All subjects received a small fee (\$10/h) or partial course credit for participating in the experiment, and were naive with respect to the task. All subjects provided informed consent prior to the start of the experiment. Seventy-four of the subjects were female; the remaining 29 were male.

2.2. Apparatus and stimuli

Stimuli were generated in Matlab (The Mathworks, v. 5.2) using the Psychophysics and Video Toolboxes (Brainard, 1997; Pelli, 1997). Stimuli were displayed on a 21" Sony Trinitron monitor at a resolution of 1024×768 pixels, which at the viewing distance of 114 cm translated to 26.7 pixels per cm. Frame rate was

85 Hz. Average luminance was 62.51 cd/m^2 . The monitor calibration data were used to build a 1779-element lookup table (Tyler, Liu, McBride, & Kontsevich, 1992) and customized computer software constructed the stimuli on each trial by selecting the appropriate luminance values from the calibrated lookup table and storing them in the display's eight-bit lookup table.

The stimuli were faces of five males and five female faces cropped to show only internal features and equated for spatial frequency content. The methods used to create the stimuli have been described previously (Gold, Bennett, & Sekuler, 1999a; Gold et al., 1999b). Stimulus size was 256×256 pixels, subtending $4.8 \times 4.8^\circ$ of visual angle from the viewing distance of 114 cm. During the experiment, stimulus contrast was varied across trials using the method of constant stimuli. The seven levels of contrast were spaced equally on a logarithmic scale, and spanned a range that was sufficient to produce significant changes in performance in virtually all subjects. The images were shown in three levels (low, medium and high) of static two-dimensional Gaussian noise, created by sampling from distributions with variances of 0.001, 0.01, and 0.1. Hence, there were a total of 21 stimulus conditions (seven contrast levels \times three external noise levels) that allowed subjects to view each face at a variety of signal-to-noise ratios.

2.3. Procedure

All subjects performed a face identification task in two sessions (Fig. 1). Subjects in the 12HR-Sleep group ($n = 24$) performed Session 1 at 9pm and Session 2 at 9am the next day. Subjects in the 12HR-No-Sleep group ($n = 24$) performed Session 1 at 9am and Session 2 at 9pm on the same day. Thus, there was a 12-hour interval between sessions in both the 12HR-Sleep and 12HR-No-Sleep groups. In the 9AM-Sleep group ($n = 24$), subjects performed Sessions 1 and 2 at 9am on Days 1 and 2. In the 9PM-Sleep group ($n = 24$), subjects performed Sessions 1 and 2 at 9pm on Days 1 and 2. Therefore, there was a 24-hour interval between sessions in both the 9am-Sleep and 9pm-Sleep groups. Subjects in the sleep groups were instructed to sleep normally overnight (between sessions), whereas subjects in the 12HR-No-Sleep group were instructed to not sleep or nap between sessions. An additional

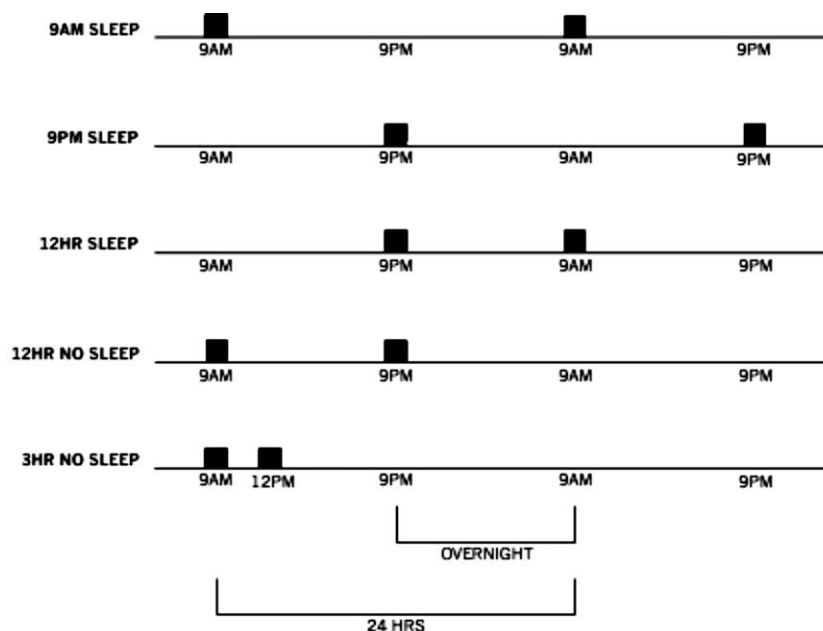


Fig. 1. Illustration of the experimental design. Placeholders indicate the time of testing for each group across two consecutive days.

group of seven subjects in the 3HR-No-Sleep was tested with a three-hour interval between sessions (Session 1 at 9am and Session 2 at noon on the same day), with instructions not to nap between sessions.

During each session, subjects were seated in a darkened room 114 cm away from the monitor. Viewing was binocular, and viewing position and distance were stabilized with an adjustable chin/forehead rest. The experiment started after a 60 s period during which the subject adapted to the average luminance of the display. A trial began with the presentation of a fixation point – a black high-contrast spot (0.15×0.15 deg) – in the center of the screen for 100 ms, followed by a randomly selected face presented for approximately 200 ms at the center of the screen. After the face disappeared, the entire set of 10 faces was presented as two rows of five noiseless, high-contrast thumbnail images, each subtending approximately $1.7 \times 1.7^\circ$. The subject’s task was to decide which one of the 10 faces had been presented during the trial by selecting one of the thumbnail images with a computer mouse. The location of each face in the response window was constant across subjects, trials, and sessions. Auditory feedback in the form of high-pitched (correct) and low-pitched (incorrect) tones informed the subject about the accuracy of each response, and the next trial began one second after presentation of the feedback. Session 1 comprised 20 trials per stimulus condition for a total of 420 trials. Session 2 comprised 40 trials per stimulus condition for a total of 840 trials. Sessions 1 and 2 lasted approximately 30 and 60 min, respectively.

Subjects completed the Morningness-Eveningness questionnaire (Horne & Ostberg, 1976), which assesses the time of day at which the subjects perform optimally. The questionnaire was completed prior to the start of the experiment in Session 1.

3. Results

Statistical analyses were done with R (R Development Core Team, 2007). All *t*-tests were two-tailed and assumed unequal group variances. Where appropriate, we report 95% confidence intervals (95%CI) for estimated parameters and Cohen’s *f* as a measure of effect size (Cohen, 1988).

Due to experimenter error, seven subjects did not complete the Morning–Eveningness questionnaire. Analysis of the completed questionnaires showed that subjects were unbiased as to their preferred time of day ($M = 46$, $SD = 9.18$), and that the preferred time of day did not vary across the five groups, $F(5,90) = 0.8275$, $p = 0.53$. Preliminary analyses also indicated that questionnaire scores were not associated with task performance, and therefore we did not include them in subsequent analyses.

Fig. 2a and b show proportion correct (collapsed across stimulus contrast and noise levels) in each session. In Session 1, the average proportion correct was 0.44 and did not vary significantly across groups, $f = 0.11$, $F(4,98) = 1.32$, $p = 0.267$, indicating that the time of day did not affect initial performance. In Session 2, the average proportion correct was 0.59, and also did not vary significantly across groups, $f = 0.18$, $F(4,98) = 1.85$, $p = 0.13$. Response accuracy

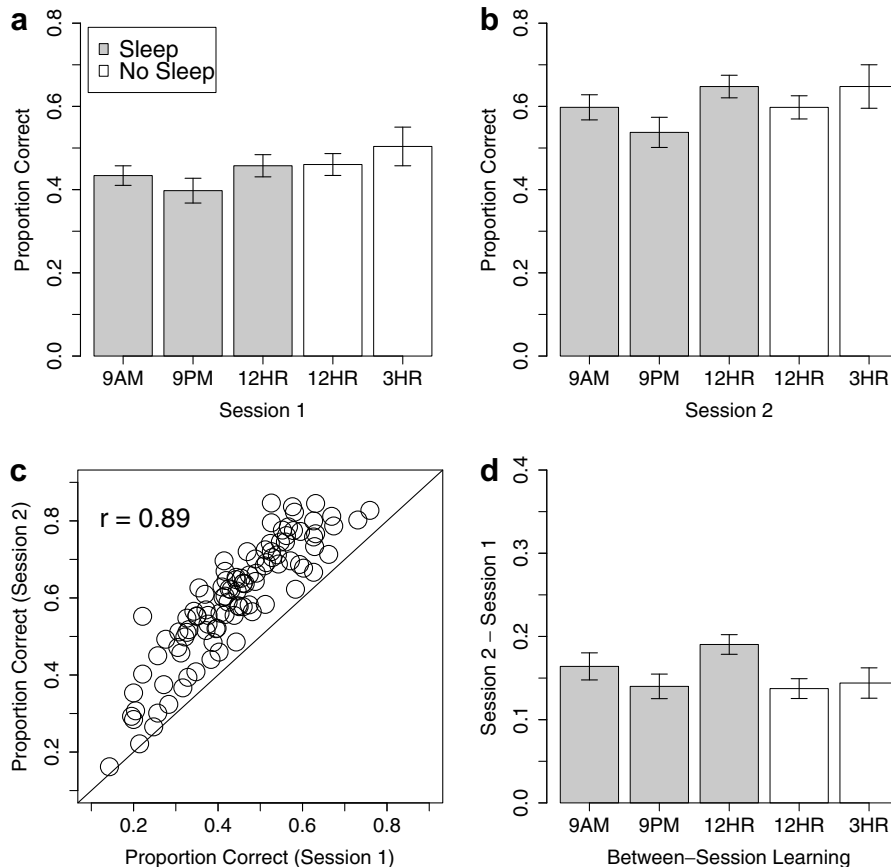


Fig. 2. (a) Proportion correct during Session 1. Performance did not vary significantly across groups. (b) Proportion correct during Session 2. Performance did not vary significantly across groups. (c) A scatter plot showing the association between proportion correct in Sessions 1 and 2. The diagonal line indicates equal performance in both sessions. Note that all points are above the line, indicating that all subjects showed some improvement across sessions. The correlation between sessions was 0.89. (d) The amount learned, defined as the difference between response accuracy in Sessions 1 and 2, for each group. There was a significant difference between groups.

in Session 1 was highly correlated with accuracy in Session 2, $r = 0.89$, $t(101) = 19.82$, $p < 0.0001$, demonstrating that our measures were reliable (Fig. 2c).

Between-session learning – defined as the difference between response accuracy in Sessions 1 and 2 – is plotted in Fig. 2d for each group. The increase in response accuracy across sessions was significant, $95\%CI = (0.144, 0.170)$, $t(102) = 23.2$, $p < 0.0001$ and was highly reliable across subjects. Indeed, every subject showed an increase in response accuracy across sessions (see Fig. 2c). There was a significant effect of Group, $f = 0.25$, $F(4,98) = 2.55$, $p = 0.044$, indicating that some groups improved more than others. t -tests were used to conduct orthogonal comparisons that evaluated the effects of sleep, interval between sessions, and time of day on the amount learned. A t -test that compared the groups that slept (9AM, 9PM, 12HR) to the groups that did not sleep (12HR, 3HR) found a marginally significant effect of sleep: the amount of learning was slightly greater in the Sleep groups than in the No-sleep groups, $95\%CI = (-0.002, 0.052)$, $t(101) = 1.97$, $p = 0.052$. A second t -test found that more learning occurred in the 12HR-Sleep group than in the two 24 h sleep groups, $95\%CI = (0.006, 0.071)$, $t(70) = 2.375$, $p = 0.021$. A third t -test that compared the 9AM-Sleep group to the 9PM-Sleep group found no effect of time-of-day on the amount learned, $95\%CI = (-0.02, 0.058)$, $t(46) = 1.093$, $p = 0.28$. These analyses suggest that learning was slightly greater in subjects that slept between experimental sessions, and that the benefits of sleep were greater when the sessions were separated by 12 h instead of 24 h.

The time-course of within-session learning was examined by measuring the proportion of correct responses that occurred in separate bins of 105 trials. This procedure yielded four bins within Session 1 and eight bins within Session 2. Fig. 3 shows proportion correct for all groups plotted as a function of bin number. During Session 1, performance improved on average by 0.18 from bin 1 to bin 4. An analysis of variance revealed a significant main effect of Bin, $f = 1.02$, $F(3, 294) = 143.43$, $p < 0.0001$, but the main effect of Group, $f = 0.05$, $F(4,98) = 1.32$, $p = 0.27$, and the Group - Bin interaction, $f = 0.04$, $F(12,294) = 1.05$, $p = 0.40$, were not significant. Further analyses showed that the linear, $F(1,98) = 256.98$, $p < 0.0001$, and quadratic, $(F(1,98) = 22.90$, $p < 0.0001)$, trends across bins were both significant, and that neither trend interacted with Group (Group \times Linear Trend: $F(4,98) = 1.02$, $p = 0.4$; Group \times Quadratic Trend: $F(4,98) = 0.53$, $p = 0.71$). During Session 2, proportion correct increased by 0.1 from bin 5 to bin 12. An analysis of variance revealed a significant main effect of Bin, $f = 0.54$, $F(7, 686) = 36.1$, $p < 0.0001$, but the main effect of Group, $f = 0.06$, $F(4,98) = 1.85$, $p = 0.13$, and the Bin \times Group interaction, $f = 0$,

$F(28, 686) = 0.89$, $p = 0.62$, were not significant. Furthermore, the linear, $F(1,98) = 78.5$, $p < 0.0001$, and quadratic, $F(1,98) = 52.4$, $p < 0.0001$, trends were both significant, and neither trend differed significantly across groups (Group \times Linear Trend: $F(4,98) = 1.48$, $p = 0.21$; Group \times Quadratic Trend: $F(4,98) = 0.74$, $p = 0.57$). We also conducted a more focussed test of the effect of sleep on Day 2 by combining all subjects in the three sleep groups into one group, and all subjects in the two no-sleep groups into another group. An ANOVA on these new groups (Sleep vs. No-sleep) found the same, significant main effect of Bin that was found in the previous ANOVA, a non-significant main effect of Sleep, $f = 0$, $F(1,101) = 0.2$, $p = 0.65$, and a non-significant Sleep \times Bin interaction, $f = 0$, $F(28,686) = 0.89$, $p = 0.62$. Hence, the overall trends were similar in Sessions 1 and 2: performance improved significantly within each session, the time-course of within-session learning did not vary across groups, and there was no indication that performance differed significantly between sleep and no-sleep groups.

A close examination of Fig. 3 suggests that the presence or absence of sleep affected performance at the start of Session 2. Specifically, performance in the No-Sleep groups was less accurate in bin 5 than bin 4 (bin5 – bin4 = -0.03), and then increased in bin 6 (bin6 – bin5 = 0.07). In the sleep groups, on the other hand, performance increased monotonically across bins 4 through 6 (bin5 – bin4 = 0.01 ; bin6 – bin5 = 0.05). A trend analysis confirmed that the quadratic trend in performance across bins 4–6 differed significantly across groups, $F(1,98) = 5.52$, $p = 0.02$. The origin of this effect is shown in Fig. 4, which plots difference scores calculated for response accuracy in bins 4 and 5. The boxplots indicate that response accuracy was not higher in bin 5 than bin 4 in a substantial proportion of subjects in each group, and that the median difference scores were close to zero for all groups. However, the difference scores were slightly lower in the No-Sleep groups. Hence, the data suggest that sleep, rather than boosting performance, prevented a slight deterioration in performance at the start of Session 2.

A global measure of within-session learning was defined as the difference between proportion correct measured in the first and last bins: for example, learning during Session 1 was the difference between proportion correct in bins 4 and 1. Surprisingly, estimates of within-session learning from Sessions 1 and 2 were not correlated, but each measure of within-session learning was correlated with between-session learning (Fig. 5). To determine if the association of within- and between-session learning varied across groups, we evaluated the interaction term in linear models of the form

$$b = w + G + wG \quad (1)$$

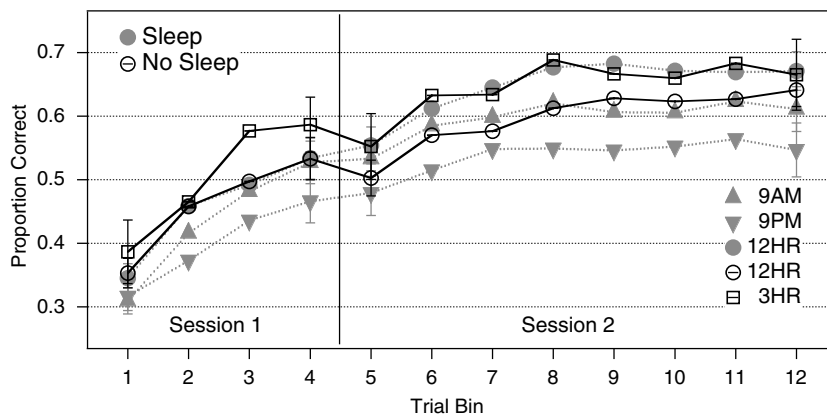


Fig. 3. Time-course of learning across both sessions for all groups. The proportion of correct responses, averaged across subjects, is plotted for bins of 105 trials. Bins 1–4 are from Session 1, whereas bins 5–12 are from Session 2. The filled and unfilled symbols represent data from Sleep and No Sleep groups, respectively. Error bars represent ± 1 standard error. For clarity, error bars are shown for four bins only. Error bars in other bins were similar to the ones shown in the figure.

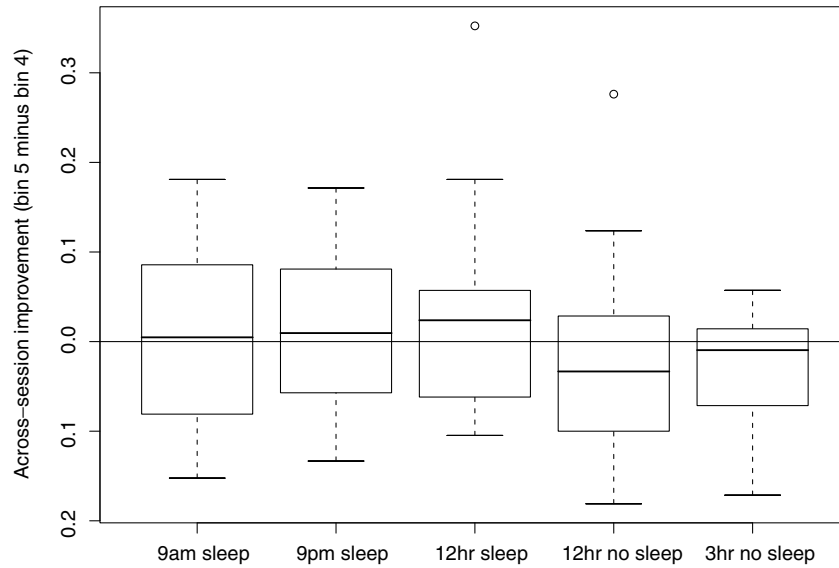


Fig. 4. Boxplots illustrating the changes in response accuracy (i.e., bin 5 – bin 4) for five groups. In each boxplot, the box depicts the interquartile range (IQR), the horizontal line indicates the group median, and the whiskers extend to the most extreme data points that are within $1.5 \times$ IQR of the median. The unfilled symbols represent outliers.

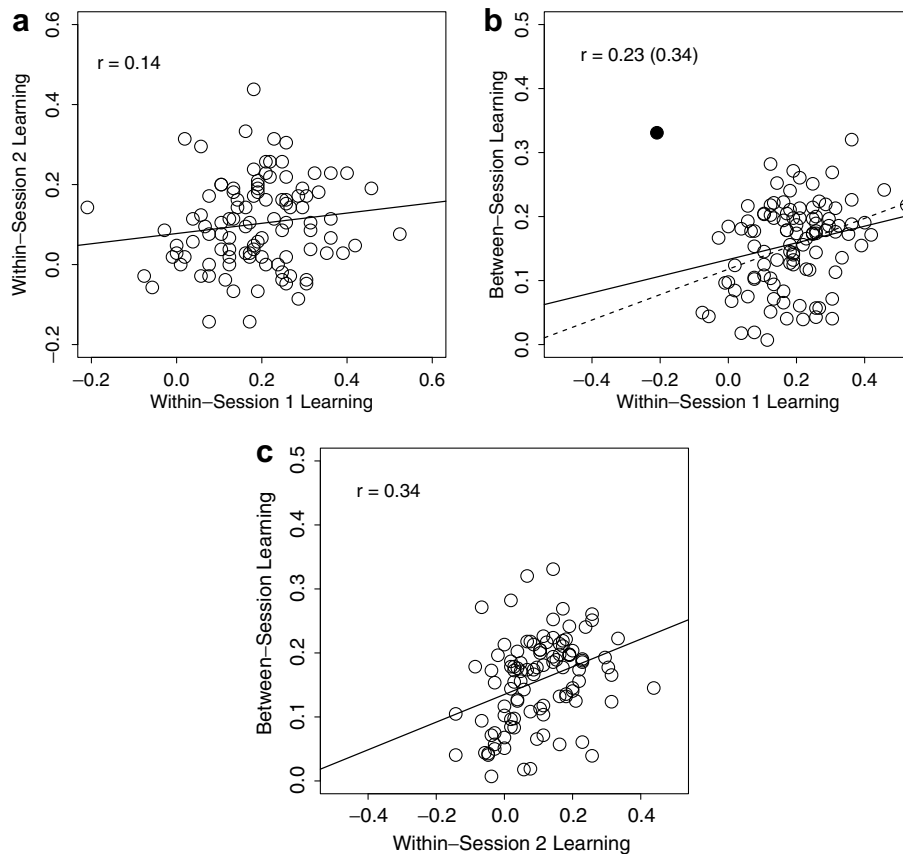


Fig. 5. Scatter plots showing the association between learning in Sessions 1 and 2. Within-session learning was defined as the difference between proportion correct in the first and last bins of a single session. Between-session learning was defined as the difference in the proportion correct between Sessions 1 and 2. Solid lines in each panel represent the best-fitting (least-squares) fit to the data. (a) Within-session learning from the two sessions was not correlated ($r = 0.14$, $t(101) = 1.42$, $p = 0.16$). (b) Within-session learning from Session 1 was significantly correlated with between-session learning ($r = 0.23$, $t(101) = 2.35$, $p = 0.02$). One subject, indicated by the solid symbol, was an outlier: excluding that subject increased the correlation to 0.34, $t(100) = 3.61$, $p = 0.0005$. The dashed line is the regression line computed after excluding the outlier. (c) Within-session learning from Session 2 was significantly correlated with between-session learning ($r = 0.34$, $t(101) = 3.71$, $p = 0.0003$).

where b is between-session learning (see Fig. 2d), w is within-session learning, G is a factor representing Group, and wG is the $w \times G$ interaction. The analyses indicated that the association

between learning during Session 1 and between-session learning did not vary across groups, $f = 0$, $F(4,93) = 0.62$, $p = 0.69$. Likewise, the association between learning during Session 2 and between-

session learning did not differ across groups, $f = 0.04$, $F(4,93) = 1.40$, $p = 0.24$. Hence, the analyses show that the magnitudes of within- and between-session learning were correlated, and that the correlation was similar across groups.

Hauptmann, Reinhart, Brandt, and Karni (2005) reported that between-session learning in a letter enumeration task was larger in subjects whose performance had reached an asymptotic value during the first session. To determine whether a similar effect held in the current experiment, we first examined how proportion correct varied across bins 1–4 in individual subjects. We identified 45 subjects whose proportion correct was approximately constant or declined in bins 2–4; response accuracy in the remaining 58 subjects increased in bins 2–4 and showed no signs of reaching an upper asymptote. These two groups of subjects were represented by a binary classification factor that indicated if a subject did or did not reach an upper limit during Session 1. Finally, this classification factor was added to a model that predicted between-session learning from within-session learning during Session 1, Group membership, and the interaction between those two predictor variables (Eq. 1). Adding the binary classification factor did not improve the model's fit significantly, $f = 0$, $F(1,92) = 0.38$, $p = 0.54$. Hence, there was no evidence that the predictability of between-session learning from learning during Session 1 differed between subjects whose performance did and did not reach an upper asymptote during the first session.

3.1. Face identification thresholds

Our experiment used a fixed set of contrasts for all subjects, and therefore was not optimized to measure thresholds accurately in individuals. Nevertheless, it was possible to fit psychometric functions to all but eight of the 618 individual sets of data. The resulting thresholds, defined as the rms contrast that corresponded to 50% correct responses, exhibited significant positive skew and contained several outliers in each condition. Therefore, we used an M-estimator of central tendency, rather than the mean, to represent the “typical” threshold in each condition (Wilcox, 2005). Fig. 6 shows the M-estimator of threshold (Huber's Ψ ; Huber, 1981) measured in each group at each noise level on both days of testing.

A percentile bootstrap procedure (Wilcox, 2005, page 310) was used to assess group differences in the M-estimator of threshold at each level of noise on each day of testing. The familywise probability of a Type I error was set to $\alpha = 0.05$ for each session. None of the bootstrap tests was significant on Session 1 or Session 2. To compare thresholds in subjects who did and did not sleep between sessions, we combined thresholds from each of the three Sleep groups into a single Sleep group, and thresholds from the two No-sleep groups into a single No-sleep group. M-estimators of thresholds in the combined groups were then compared at each level of noise. Again, none of the comparisons was significant on either day. In summary, we did not obtain clear evidence that thresholds varied across the experimental groups, or between subjects that did or did not sleep between sessions, in Session 1 or 2.

The log-difference between thresholds in Sessions 1 and 2 was calculated for each subject in each condition. The M-estimator of the log-difference between thresholds was -0.27 , -0.62 , and -0.56 in the low, medium, and high external noise conditions, respectively. All three of these M-estimators differed significantly from zero ($p < 0.05$), which shows that thresholds decreased significantly at all levels of noise. To determine if the decrease in threshold varied across groups, a percentile bootstrap was used to compare the M-estimators measured for each group at each level of noise. None of the bootstrap tests was significant ($p > 0.05$). Next, the log-difference scores for subjects in the three Sleep groups were combined into a single Sleep group, and the scores from the remaining groups were combined into a single No-sleep

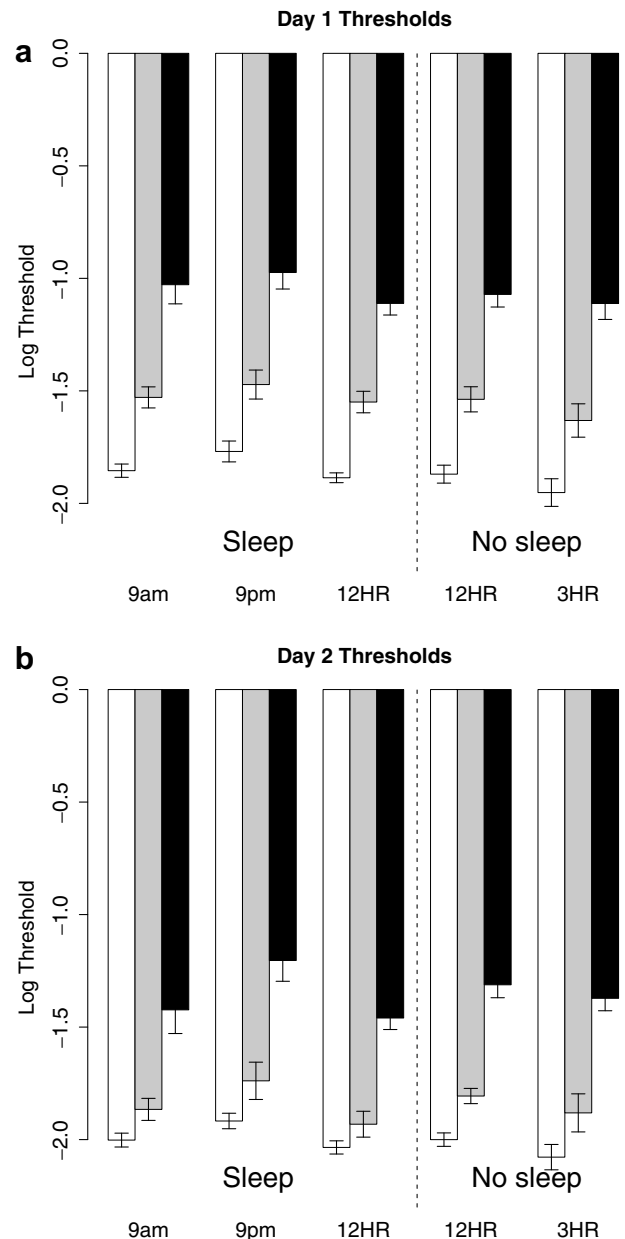


Fig. 6. Log-transformed threshold, expressed in terms of rms contrast, for each group in each condition on Day 1 (a) and Day 2 (b) of testing. Each bar represents the M-estimator of threshold measured with low (unfilled bars), medium (grey bars), or high (black bars) levels of external noise. M-estimators are robust measures of central tendency: in cases where the data are skewed, as was the case in the current experiment, an M-estimator is better than the mean at providing an estimate of a “typical” value (Wilcox, 2005). The particular M-estimator used here is Huber's Ψ . The error bars represent ± 1 standard error of the M-estimator.

group. Again, a percentile bootstrap performed on the M-estimators of the log-difference between thresholds found no significant ($p > 0.05$) difference between the combined Sleep and No-sleep groups at any level of external noise. These analyses indicate that the difference between thresholds measured in the two sessions did not vary significantly across groups or between subjects that did or did not sleep.

In summary, we did not obtain clear evidence that thresholds in Sessions 1 and 2, or the difference between thresholds in the two sessions, varied across the experimental groups or was associated with the presence or absence of sleep. In this regard, these analyses of thresholds are consistent with the previous analyses on overall proportion correct.

4. Discussion

Using a 1-of-10 face identification task, we found that the proportion of correct responses increased by 0.17 and 0.1 during the first and second testing sessions, respectively. The difference between overall response accuracy in the two sessions was approximately 0.15, and therefore the magnitude of within-session and between-session learning effects was nearly the same. Between-session learning was significantly higher in groups that slept in-between sessions, but the effect of sleep on proportion correct was small (i.e., ≈ 0.026) and was restricted to performance in the first 105 trials in Session 2.

The finding that sleep had very small effects on learning differs from results obtained in several studies that used the texture discrimination task (e.g., Karni & Sagi, 1991; Karni et al., 1994; Karni & Sagi, 1993). Karni and Sagi (1991), for example, found that thresholds declined by approximately 22% between the first and second days of testing (see Fig. 2 in Karni & Sagi, 1991). Censor, Karni, and Sagi (2006) found that the size of sleep-related improvement in the texture discrimination task was a non-monotonic function of the number of trials used in the first session, with 26 trials per block (3–4 blocks per testing session) producing more overnight learning than 50 and 12 trials per block. Based on Censor et al.'s findings, it is tempting to attribute the small sleep effect found in the current experiment to the relatively large number of trials (i.e., 420) used in Session 1. However, using the same task and methods as in the current experiment, Hussain, Bennett, and Sekuler (2003) found that reducing the number of practice trials in Session 1 does not result in greater amounts of between-session learning: during Session 2, subjects who received 420–840 trials during Session 1 performed significantly better than subjects who received 21–210 trials during Session 1. Therefore we think it is unlikely that reducing the number of trials in Session 1 would increase the effect of sleep.

Mednick et al. (2002) reported that repeated, within-day testing on the texture discrimination task resulted in progressively higher thresholds (also see Mednick et al., 2003, 2005; but see Fig. 2 in Stickgold et al., 2000a for a different result obtained with similar procedures). Between Sessions 1 and 2, for example, thresholds increased by 17% in Mednick et al. (2002) and 15% in Mednick et al. (2005). Some studies have also reported that performance deteriorates within a single session (Mednick et al., 2005; Ofen, Moran, & Sagi, 2007). Ofen et al., for example, found that performance with a single, above-threshold stimulus in the texture discrimination task decreased from 90% correct to 70% during the course of eight blocks in a single test session (see Fig. 4, in Ofen et al., 2007). In this experiment, both No-Sleep groups did exhibit a drop in performance at the start of Session 2, but the effect was small and was restricted to the first bin of 105 trials. Instead of decreasing, overall accuracy increased across sessions in both No-Sleep groups (see Fig. 2d). Furthermore, we found no evidence of performance decreasing within a session; in fact, response accuracy increased within each session in all groups.

This experiment differs in several ways from previous studies that found larger effects of sleep or within-day deterioration. One potentially important difference concerns the psychophysical methods used to assess performance. Previous studies that measured learning with the texture discrimination task have obtained thresholds by adjusting the stimulus level using a variation of the traditional descending method of limits. With this method, increasing the number of trials per block, as was done by Censor et al. (2006) and Ofen et al. (2007), would necessarily reduce the variety of signal-to-noise ratios presented to the subject over the course of several trials or minutes. In contrast, this experiment presented stimuli that varied significantly in terms of signal-to-noise

ratio in a random order. This randomization procedure may minimize the adaptation that is thought to be important for generating within-session deterioration (Ofen et al., 2007).

Another obvious difference between studies is that this experiment used a face identification task rather than the texture discrimination task. Face identification may depend on higher-level mechanisms that differ significantly from the mechanisms tapped by the texture discrimination task used in previous studies. It is possible, therefore, that learning in a face identification task differs qualitatively from the learning that is found with simpler perceptual tasks. However, results from other experiments are inconsistent with this hypothesis. For example, learning in a face identification task is specific for both the orientation and identity of the trained items (i.e., there is little generalization to new stimuli, Hussain et al., 2005, Hussain, Bennett, & Sekuler, 2006). Furthermore, the benefits of face identification learning are long lasting, persisting for at least 9–18 months after the training sessions have ended (Hussain, Bennett, & Sekuler, 2007). Finally, face identification learning exhibits a fast within-session component and a slow between-session component (Fig. 3), as has been found with simpler tasks (e.g., Karni & Sagi, 1993). Hence, the available evidence indicates that learning the face identification task exhibits many of the characteristics of learning found with simpler perceptual tasks.

Nevertheless, the relative complexity of the neural network underlying face processing may alter the effects of sleep on learning. Sleep-dependent consolidation is thought to require the reactivation, during sleep, of cells engaged during the task. For example, consolidation of spatial learning has been linked to the reactivation of cells in the hippocampus (Wilson & McNaughton, 1994), and consolidation of visual conditioning is contingent on reactivation of visual cortex (Amzica, Neckelmann, & Steriade, 1997). Indeed, some suggest that cortical activity in V1 is required during sleep for learning-related plasticity to be enabled (Jha et al., 2005). For the texture task, an imaging study indicates that V1 is the locus of practice effects (Schwartz, Maquet, & Frith, 2002), and the consolidation of such learning may depend on the reactivation of V1 during sleep. The learning we find with more complex stimuli may involve more than one cortical locus (such as IT), possibly diluting the effects of sleep if the entire network engaged during training is not reactivated during sleep.

Our results are consistent with the sustained improvements found in the absence of sleep for auditory and motor tasks, with mere passage of time after training (Gottselig et al., 2004; Robertson et al., 2004; Roth et al., 2005). In Roth et al. (2005), improvements on a verbal identification task emerged after at least 6 h had elapsed. In Robertson et al. (2004), there was improvement on a finger-tapping task 12 h after practice with no sleep between sessions. In Gottselig et al. (2004), restful waking, but not busy waking was equivalent to sleep in promoting learning of an auditory pattern discrimination task, suggesting that the key to consolidation might be the absence of interference from other tasks during the interim period (i.e., the interference hypothesis). However, performance on this auditory task also improved with no break between sessions, suggesting that the benefits from consolidation may be superfluous to those gained from continuous task performance. Likewise, face-identification requires little latent processing, as is clear from the performance of the 3 h-No sleep group, which improved despite the small time-window between sessions. These results are also consistent with a recent study showing negligible effects of sleep in face memory task (Sheth, Nguyen, & Janvelyan, 2008).

We found that within- and between-session learning were correlated positively for all groups (Fig. 5). This finding is at odds with the results of Walker et al. (2003), who reported that sleep-dependent and within-session improvements on a motor task were

uncorrelated. It should be pointed out, however, that although we found that the correlation was significant, within-session learning accounted for only a small proportion of the variance in between-session learning (i.e. $r^2 = 0.34^2 = 0.11$). It is unlikely that this relatively weak association is due to low reliability of our dependent measure, because the correlation between performance in Session 1 and 2 was 0.89 (Fig. 2c). Instead, it appears that between-session learning depends substantially on factors that are not correlated with within-session learning. Stated this way, these findings are similar to those reported by Walker et al. but, in addition, suggest that their conclusion that within- and between-session learning are mediated by “distinct and independent processes” (page 281, Walker et al., 2003) is too strong.

Overall, our results show that robust perceptual learning for a face identification task can be obtained in the absence of sleep, and that sleep has very little effect on between-session learning. The time-course of learning within each session was also unaffected by the presence or absence of sleep between sessions. These results are inconsistent with the notion that the bulk of perceptual learning is a latent, across-session phenomenon. It is commonly assumed that the latent gains are the basis of stimulus-specificity typically found in perceptual learning studies (Karni & Bertini, 1997; Karni et al., 1998; Karni & Sagi, 1993). The present findings, combined with other results from our lab indicating exemplar- and orientation-specificity of face learning (Hussain et al., 2005, 2006), suggest that stimulus-specificity of learning could just as well emerge from the improvements that occur within the training session.

References

- Amzica, F., Neckelmann, D., & Steriade, M. (1997). Instrumental conditioning of fast (20- to 50-Hz) oscillations in corticothalamic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 94(5), 1985–1989.
- Atienza, M., Cantero, J. L., & Stickgold, R. (2004). Posttraining sleep enhances automaticity in perceptual discrimination. *Journal of Cognitive Neuroscience*, 16(1), 53–64.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Brashers-Krug, T., Shadmehr, R., & Bizzi, E. (1996). Consolidation in human motor memory. *Nature*, 382(6588), 252–255.
- Censor, N., Karni, A., & Sagi, D. (2006). A link between perceptual learning, adaptation and sleep. *Vision Research*, 46(23), 4071–4074.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: L. Erlbaum Associates.
- Fahle, M. (1994). Human pattern recognition: Parallel processing and perceptual learning. *Perception*, 23(4), 411–427.
- Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, 35(21), 3003–3013.
- Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature*, 425(6958), 614–616.
- Fiorntini, A., & Berardi, N. (1981). Learning in grating waveform discrimination: Specificity for orientation and spatial frequency. *Vision Research*, 21(7), 1149–1158.
- Fischer, S., Hallschmid, M., Elsner, A. L., & Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences of the United States of America*, 99(18), 11987–11991.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999a). Identification of band-pass filtered letters and faces by human and ideal observers. *Vision Research*, 39(21), 3537–3560.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999b). Signal but not noise changes with perceptual learning. *Nature*, 402(6758), 176–178.
- Gottselig, J. M., Hofer-Tinguely, G., Borbely, A. A., Regel, S. J., Landolt, H.-P., Retey, J. V., et al. (2004). Sleep and rest facilitate auditory learning. *Neuroscience*, 127(3), 557–561.
- Hauptmann, B., Reinhart, E., Brandt, S. A., & Karni, A. (2005). The predictive value of the leveling off of within session performance for procedural memory consolidation. *Brain Research. Cognitive Brain Research*, 24(2), 181–189.
- Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology*, 4(2), 97–110.
- Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
- Hussain, Z., Bennett, P. J., & Sekuler, A. B. (2003). How much practice is needed to produce perceptual learning? *Journal of Vision*, 3(9), 256a.
- Hussain, Z., Bennett, P., & Sekuler, A. (2005). Perceptual learning of faces and textures is tuned to trained identities. *Perception: ECVF Abstract Supplement*, 34.
- Hussain, Z., Bennett, P. J., & Sekuler, A. B. (2006). Face-inversion effects flex with perceptual learning. *Journal of Vision*, 6(6), 153a.
- Hussain, Z., Bennett, P., & Sekuler, A. (2007). Superior identification of familiar visual patterns a year after learning. *Journal of Vision*, 7(9), 797a.
- Jha, S. K., Jones, B. E., Coleman, T., Steinmetz, N., Law, C.-T., Griffin, G., et al. (2005). Sleep-dependent plasticity requires cortical activity. *Journal of Neuroscience*, 25(40), 9266–9274.
- Karni, A., & Bertini, G. (1997). Learning perceptual skills: behavioral probes into adult cortical plasticity. *Current Opinion in Neurobiology*, 7(4), 530–535.
- Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., et al. (1998). The acquisition of skilled motor performance: fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 95(3), 861–868.
- Karni, 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(11), 4966–4970.
- Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443), 250–252.
- Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., & Sagi, D. (1994). Dependence on rem sleep of overnight improvement of a perceptual skill. *Science*, 265(5172), 679–682.
- McKee, S. P., & Westheimer, G. (1978). Improvement in vernier acuity with practice. *Perception and Psychophysics*, 24(3), 258–262.
- 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(10), 3881–3885.
- Mednick, S. C., Nakayama, K., Cantero, J. L., Atienza, M., Levin, A. A., Pathak, N., et al. (2002). The restorative effect of naps on perceptual deterioration. *Nature Neuroscience*, 5(7), 677–681.
- Mednick, S., Nakayama, K., & Stickgold, R. (2003). Sleep-dependent learning: a nap is as good as a night. *Nature Neuroscience*, 6(7), 697–698.
- Ofen, N., Moran, A., & Sagi, D. (2007). Effects of trial repetition in texture discrimination. *Vision Research*, 47(8), 1094–1102.
- Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, 256(5059), 1018–1021.
- Press, D. Z., Casement, M. D., Pascual-Leone, A., & Robertson, E. M. (2005). The time course of off-line motor sequence learning. *Brain Research. Cognitive Brain Research*, 25(1), 375–378.
- R Development Core Team, 2007. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available from (<http://www.R-project.org>).
- Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. *Current Biology*, 14(3), 208–212.
- Roth, D. A.-E., Kishon-Rabin, L., Hildesheimer, M., & Karni, A. (2005). A latent consolidation phase in auditory identification learning: time in the awake state is sufficient. *Learning & Memory*, 12(2), 159–164.
- Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: retinotopy, orientation specificity and monocularity. *Journal of Physiology*, 483(Pt 3), 797–810.
- Schwartz, S., Maquet, P., & Frith, C. (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(26), 17137–17142.
- Sheth, B., Nguyen, N., & Janvelyan, D. (2008). How sleep influences our memory for faces. *Journal of Vision*, 8(6), 876a.
- Siegel, J. M. (2001). The rem sleep-memory consolidation hypothesis. *Science*, 294(5544), 1058–1063.
- Song, S., Howard, J. H. J., & Howard, D. V. (2007). Sleep does not benefit probabilistic motor sequence learning. *Journal of Neuroscience*, 27(46), 12475–12483.
- Stickgold, R., James, L., & Hobson, J. A. (2000a). Visual discrimination learning requires sleep after training. *Nature Neuroscience*, 3(12), 1237–1238.
- Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., & Hobson, J. A. (2000b). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience*, 12(2), 246–254.
- Tyler, C.W., Liu, H.C.L., McBride, B., & Kontsevich, L. (1992). Bit-stealing: How to get 1786 or more gray levels for an 8-bit color monitor. In: Rogowitz, B.E., (Ed.), *Proceedings of SPIE: Human vision, visual processing and digital display III*. (Vol. 1666, pp. 351–364).
- Vertes, R. P. (2004). Memory consolidation in sleep; dream or reality. *Neuron*, 44(1), 135–148.
- Vertes, R. P., & Eastman, K. E. (2000). The case against memory consolidation in REM sleep. *The Behavioral and Brain Sciences*, 23(6), 867–876.
- Walker, M. P., Brakefield, T., Seidman, J., Morgan, A., Hobson, J. A., & Stickgold, R. (2003). Sleep and the time course of motor skill learning. *Learning & Memory*, 10(4), 275–284.
- Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, 44(1), 121–133.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd edition). Amsterdam: Elsevier/Academic Press.
- Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Journal of Science*, 265(5172), 676–679.