AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology:
Learning, Memory, and Cognition

© 2022 American Psychological Association
ISSN: 0278-7393

https://doi.org/10.1037/xlm0001090

Severe Publication Bias Contributes to Illusory Sleep Consolidation in the Motor Sequence Learning Literature

Timothy C. Rickard[1], Steven C. Pan[2], and Mohan W. Gupta[1]
[1] Department of Psychology, University of California, San Diego
[2] Department of Psychology, University of California, Los Angeles

We explored the possibility of publication bias in the sleep and explicit motor sequence learning literature by applying precision effect test (PET) and precision effect test with standard errors (PEESE) weighted regression analyses to the 88 effect sizes from a recent comprehensive literature review (Pan & Rickard, 2015). Basic PET analysis indicated pronounced publication bias; that is, the effect sizes were strongly predicted by their standard error. When variables that have previously been shown to both moderate the sleep gain effect and substantially reduce unaccounted for effect size heterogeneity were included in that analysis, evidence for publication bias remained strong. The estimated postsleep gain was negative, suggesting forgetting rather than facilitation, and it was statistically indistinguishable from the estimated postwake gain. In a qualitative review of a smaller group of more recent studies we observed that (a) small sample sizes—a major factor behind the publication bias—are still the norm, (b) use of demonstrably flawed experimental design and analysis remains prevalent, and (c) when authors conclude in favor of sleep-dependent consolidation, they frequently do not cite the articles in which those methodological flaws have been demonstrated. We conclude that there is substantial publication bias, that there is no consolidation-based, absolute performance gain following sleep, and that strong conclusions regarding the hypothesis of less forgetting after sleep than after wakefulness should await further research. Recommendations are made for reducing publication bias in future work.

Keywords: publication bias, sleep, motor sequence learning, finger-tapping, finger–thumb

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Over the last 2 decades, there has been substantial research interest into the role that sleep may play in consolidation of motor learning, and in particular explicit motor sequence learning. The typical experiment in that literature (e.g., Walker et al., 2002) involves either a keyboard finger-tapping task or an analogous finger-thumb opposition task wherein participants learn to type or tap a repeating sequence. Participants are given multiple training blocks to acquire proficiency on the task (most commonly, 12 blocks), with the duration of each block varying across experiments (most typically, 30 s), and with a rest period between each block (typically also of 30 s). After training, there is a delay involving either sleep (in the form of a nap or a full night) or only wakefulness. On a subsequent test, there are two or more blocks of the same task, again interleaved with rest periods. Across that literature, time of training and time of testing vary between morning and late evening, and the delay period between those sessions varies between about one hour (for studies involving naps) and 72 hr. The effect of the delay between sessions is measured by comparing performance averaged over the last few training blocks (the pretest) to performance averaged over the first few test blocks (the *posttest*).

Early studies using that experimental paradigm (e.g., Korman et al., 2003; Walker et al., 2002; Walker et al., 2003) yielded a striking result: performance improved substantially between training and test sessions (by about 20%) if sleep occurred between those sessions (the postsleep gain), but did not improve, or improved less, if only wakefulness occurred between sessions (constituting the relative sleep gain effect). By 2008, at least a dozen articles had been published showing one or both of those two effects, jointly yielding more than 6,300 citations to date. A multitude of similar articles have since been published. Almost ubiquitously in those articles, observed sleep gain effects have been interpreted to reflect a sleep-dependent consolidation process.

In several more-recent articles, however, evidence has been advanced that the empirical postsleep gain effect does not reflect a consolidation process, but rather is an artifact of several experimental and data analytic confounds. In an early example, Rickard et al. (2008) demonstrated that the postsleep gain is eliminated when (a) the duration of training blocks between breaks is reduced from the typical 30 s to 10 s, minimizing the accumulation of block-level reactive inhibition as well as more general task fatigue, two factors can inflate the postsleep gain estimate; (b) a 24-hr delay is used, thus holding time of training and testing constant

Mohan W. Gupta https://orcid.org/0000-0001-8632-592X

We thank Mark Appelbaum for providing valuable comments on a draft of the article.

The data underlying this article are available on the Open Science Framework (https://osf.io/tv3p4/).

Correspondence concerning this article should be addressed to Timothy C. Rickard, Department of Psychology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0109, United States. Email: trickard@ucsd.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

and eliminating possible time-of-day confounds (e.g., effects of circadian rhythms on performance); and (c) the postsleep gain effect is assessed using a learning curve continuity analysis rather than the nearly ubiquitous pretest-posttest difference score which can inflate the postsleep gain estimate due to averaging over online learning. Other authors have since reached similar conclusions using related approaches to minimize confounds (Brawn et al., 2010; Landry et al., 2016; Nettersheim et al., 2015), and related findings have been reported for the case of implicit motor sequence learning (Nemeth et al., 2010; Simor et al., 2018). In a comprehensive meta-analysis that included 88 effect sizes drawn from 34 studies, Pan and Rickard (2015) demonstrated that the earlier empirical results indicating no postsleep gain (e.g., Rickard et al., 2008) are evident the literature when confounding experimental design and analysis factors are statistically adjusted for.

Although that work speaks against the hypothesis of a sleep consolidation process that enhances learning, the issue remains controversial. Whereas in some more recent studies the foregoing evidence is acknowledged (Borragán et al., 2015; Humiston & Wamsley, 2018; Maier et al., 2017), most authors still link the empirical postsleep gain effect to sleep-dependent consolidation (Astill et al., 2014; Bottary et al., 2016; Breton & Robertson, 2017; Diekelmann, 2017; Gregory et al., 2014; Tucker et al., 2016; Wamsley et al., 2016).

Either in addition to or in the absence of a postsleep gain, the relative sleep gain effect is sometimes observed, and it is interpreted in the literature as evidence of more consolidation during sleep than during wakefulness. However, as pointed out in the 2015 meta-analysis (Pan & Rickard, 2015; see Figure 5 of that article), the relative gain effect has been consistently observed in only one of four experimental designs (namely, a varied time design involving a p.m. training—sleep—a.m. testing group and a p.m. training—wakefulness—a.m. testing group). Further, in the Pan and Rickard (2015), that design was demonstrated to be vulnerable to a potential time of testing confound. Yet, many researchers continue to interpret the relative gain effect, when observed, as unambiguous evidence for a sleep-dependent consolidation process (Adi-Japha & Karni, 2016; Bottary et al., 2016; Breton & Robertson, 2017; Humiston & Wamsley, 2018; King et al., 2017).

Publication Bias, Sleep, and Motor Learning

Publication bias occurs when studies that produce statistical evidence for a hypothesized phenomenon are more likely to be published than those that do not. Several factors underlying publication bias have been discussed in the literature (Ioannidis, 2005; Open Source Collaboration, 2015; Simmons et al., 2011), chiefly among them for current purposes are (a) *publication criteria*, such as a cutoff *p* value, that block publication of nonsignificant results, and (b) *reporting bias*, wherein researchers decide not to pursue publication when the observed effect is null or in the opposite direction of that expected in the literature. In neither of those cases is the unpublished study necessarily flawed. Rather, the study results may reflect natural sampling variability. The expected consequence of publication bias for a literature is an inflated average effect size in the expected direction.

The likelihood of publication bias in the literature is largely determined by the relation between the effect size and the standard error of the effect size. If there is a true effect, and if the standard error is small relative to that effect (if statistical power is high), then most studies will yield statistically significant results, facilitating publication. Relatively few studies will remain unpublished, and the estimated effect size in the literature can approximate the true effect size. If, in contrast, *SE*s are large relative to the true effect size (i.e., power is low), then fewer studies will yield statistically significant results, and publication bias can be substantial.

An important characteristic of the sleep and motor sequence learning literature that is relevant to publication bias—but that has received no prior attention—is the generally small participant sample sizes. The distribution of those sample sizes, based on the 88 groups that were included in the 2015 meta-analysis, is shown in Figure 1. For most of that literature, the sample size is small (*Mdn* = 15 per group), relative to both psychological research literature at large (*Mdn* = 40) and current recommendations for best practice (Asendorpf et al., 2013; Button et al., 2013). Because sample size is a major determinant of standard error, there appears to be substantial risk of publication bias in the sleep and motor learning literature. We explore that possibility in the current article. Specifically, we estimate the effect of publication bias using both graphical analysis and the precision effect test and precision effect test with standard errors (PET–PEESE) regression method (Stanley, 2017; Stanley & Doucouliagos, 2014). We conduct PET–PEESE tests without and with inclusion of previously identified moderating variables that have been shown to account for the bulk of the effect size heterogeneity in sleep and motor learning literature (Pan & Rickard, 2015).

Method

Meta-Analytic Procedures

If there is publication bias in the sleep and motor learning literature, then studies with large standard errors and small effect sizes, or with statistically significant effects in the unexpected direction, are the most likely to be absent from the literature. If a positive effect is expected in a literature (as in the current case for both postsleep and relative gain), then publication bias will manifest as a positive slope in a plot of the sample effect sizes (Cohen's *d* in the current case; *y*-axis) against a measure of statistical error, such as the standard errors of the effect size estimates (*SE*; *x*-axis). If there is no publication bias, then that slope is expected to be zero. PET–PEESE analysis captures that effect quantitatively using weighted least squares (WLS) regression of either *d* on standard error (PET) or *d* on the sampling variability (SE^2; PEESE). In both cases, the weighting variable is $1/SE^2$. The basic equation for PET is,

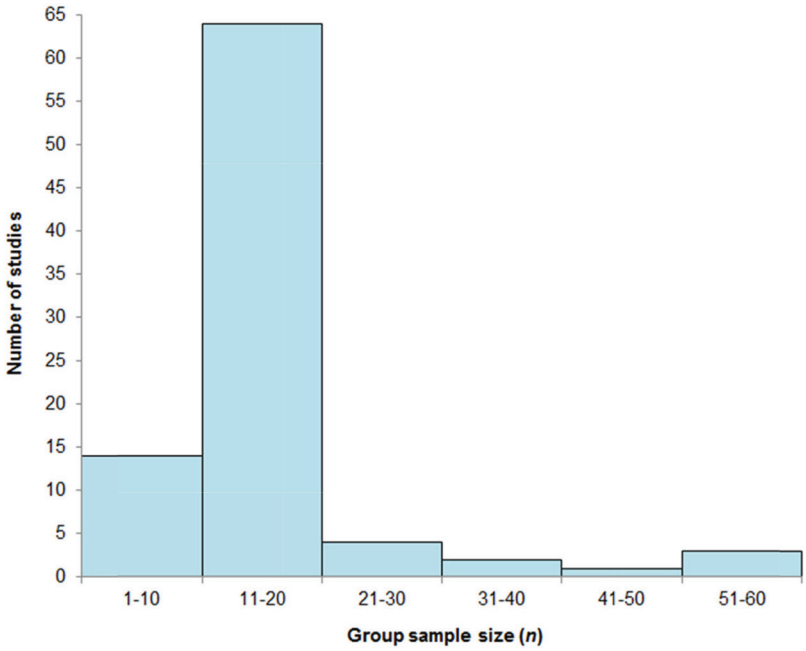
$$d = \beta_0 + \beta_1 SE_i + \varepsilon_i, \quad (1)$$

and for PEESE is,

$$d = \beta_0 + \beta_1 SE^2{}_i + \varepsilon_i, \quad (2)$$

where β_1 is the slope estimate (measuring the severity of publication bias; a slope of zero indicates no bias), SE_i is the standard error for the *i*th effect size, $SE^2{}_i$ is the corresponding sampling variability, and ε_i is the residual error. The parameter β_0 is the estimated true effect size after adjusting for publication bias (i.e., for a hypothetical experiment with a very large sample size and hence negligible sampling variability).

Figure 1
Distribution of Experimental Group Sample Sizes in the Published Sleep and Motor Sequence Literature



Note. See the online article for the color version of this figure.

PET–PEESE analysis involves three steps (in some articles, those three steps are referred to as FAT-PET–PEESE, although the first two steps involve inference based on the PET equation). First, the PET equation is fitted to the data and a significance test (at $\alpha = .05$) is performed on the β_1 estimate. If that test rejects the null, then publication bias is inferred. In that case, a significance test is performed on β_0. If that test fails to reject the null hypothesis, then there is no compelling evidence that the true effect size is different from zero after correcting for publication bias, and the procedure is terminated. If that test does reject the null, then the true effect after adjusting for publication bias is best estimated by the parameter β_0 in a PEESE fit (Equation 2).

If candidate moderating variables are included in the regression (Stanley & Doucouliagos, 2014), then the respective equations are,

$$d = \beta_0 + \beta_1 SE_i + \Sigma_k \alpha_k z_k + \varepsilon_i, \quad (3)$$

and for PEESE,

$$d = \beta_0 + \beta_1 SE^2{}_i + \Sigma_k \alpha_k z_k + \varepsilon_i, \quad (4)$$

where z_k is the k^{th} moderator variable and α_k is the corresponding effect size estimate. The three PET–PEESE steps are identical to those for the case of no moderating variables.

Recent simulation work (Alinaghi & Reed, 2018; Stanley, 2017) has shown that the PET–PEESE method yields valid results when there is effect size *homogeneity*, such that all sample effect sizes are random deviates from a *single* true effect size (i.e., the *fixed effect* case). However, in the more likely case of effect size *heterogeneity*, wherein subsets of the effect sizes are random deviations from *different* true effect sizes (i.e., the *random effects* case), results using Equations 1 and 2 can be biased. In their hierarchical random effects (HRE) meta-analysis, Pan and Rickard (2015) observed substantial heterogeneity—both between articles and over effect sizes within article. Therefore, application of Equations 1 or 2, although a useful starting point, may not yield trustworthy results regarding publication bias.

However, the 2015 meta-analysis also identified seven variables that moderated the postdelay gain effect (see Table 1). When all of those variables were simultaneously included in an HRE "final working model," the between-article heterogeneity was substantially reduced (from $\tau^2 = .27$ to .08), and the within-article heterogeneity (ω^2) was eliminated (falling from .13 to zero). Hence, by applying Equation 3 and 4 with those moderating variables included, the potential for an unaccounted-for heterogeneity influence on the PET–PEESE publication bias estimates should be greatly reduced.

Based on recent simulations (Alinaghi & Reed, 2018), two other aspects of the current dataset favor valid inference using PET–PEESE. First, when unaccounted-for within-article heterogeneity is reduced to zero, the potential bias in the PET–PEESE estimates is reduced, even if there is some unaccounted-for heterogeneity at the between-article level. In the HRE analysis noted above, inclusion of the moderators did in fact decrease within article heterogeneity to zero. Second, PET–PEESE estimation bias is further reduced when publication bias, if present, involves suppression of nonsignificant results rather than suppression of statistically significant, but negative (relative to expectation) results. Given the multiple confounding experimental design factors have been shown to cumulatively inflate postsleep gain

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 1
Effect Size Moderators for the Sleep and Motor Sequence Literature

Moderator	Description
Sleep status	Wake group vs. sleep group
Data averaging	Amount of data that was averaged to calculate pre–post performance gains between the training and test sessions
Performance duration	Amount of time per training block
Training duration	Total time devoted to training
Time of testing	Time of day that the test session occurred
Time of testing squared	Time of day that the test session occurred, squared
Elderly status	>59 years of age

Note. Moderators drawn from the final working model of the 2015 meta-analysis (Pan & Rickard, 2015).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

estimates (data averaging, duration of training, duration of each training block, Pan & Rickard, 2015), the latter type of publication bias appears to be much less likely than the former.

Data

Data used in the primary analyses correspond exactly to the data analyzed in the Pan and Rickard (2015) meta-analysis (see their Table 1). That dataset, which is publicly archived on the Open Science Framework (https://osf.io/u2c8s), encompasses 88 experimental groups (88 effect sizes) from 34 studies. Sixty-six of those were sleep groups (where sleep intervened between training and test sessions) and the remaining 23 were wake groups. Those data are comprehensive of that literature at the time, after well-justified exclusion criteria were applied. In a later section of this article, we also review the smaller, more recent literature.

Data Analysis

Data analysis was conducted in three steps. First, we compared WLS analysis with the previously conducted HRE analyses with all moderator variables included, but without the publication bias estimator, standard error The goal in that comparison was to determine whether the moderators can capture the effect size heterogeneity in the context of the WLS analysis to roughly the same extent that they do in the HRE analysis. If they can, then the current PET–PEESE analysis is well motivated. Second, we conducted PET analysis for three cases: inclusion of no moderator variables, inclusion of only the sleep status moderator, and inclusion of all moderators listed in Table 1. Third, given that the Case 3 PET analysis yielded a statistically significant (nonzero) intercept, we performed PEESE analysis for Case 3 to find the best estimate of that intercept (Step 3 of the PET–PEESE analysis described earlier).

Setting the Case 3 Intercept to Correspond to the Minimally Confounding Values of the Moderating Variables

In the most critical, Case 3 PET and PEESE analysis described below, the intercept value, β_0, is intended to estimate the effect size when the various moderator variables take values that introduce minimal confounding effects, based on the findings of Pan and Rickard (2015). Thus, the value of each variable that is minimally confounding should take a value of zero (i.e., the intercept value). For one of the continuous variables, *data averaging,* the minimally confounding intercept value corresponds to the natural intercept of zero (no averaging), so no adjustment to that variable was required. Note that analysis with zero data averaging of the training and test session data is practically achievable using a straightforward curve fitting test (Pan & Rickard, 2015). For two other continuous variables, the least confounding value as measured in the experiment was not zero, but rather a positive value. For those variables, the minimally confounding intercept was obtained by subtracting a constant value from the variable for each of the 88 effects. For p*erformance duration* per block, 10 s is both the minimum value in the literature and the least confounding value in the literature. Thus, 10 s was subtracted from the p*erformance duration* variable for each of the 88 effects. A *training duration* of 360 s is by far the most common in the literature. By the end of training with that task duration, task performance improvements over blocks is minimal, virtually negating any potential online learning confound in the pretest calculation. Hence, 360 s was subtracted from the training duration for each effect size. The dichotomous *sleep status* variable was set to a value of zero for sleep groups and 1 for wake groups. Finally, the dichotomous *elderly status* moderator was set to zero for groups with nonelderly participants and to one for groups with elderly participants, so that the intercept in Case 3 represents the postsleep gain for nonelderly participants. Elderly groups often exhibit little evidence for an immediate postsleep gain (Backhaus et al., 2016; Tucker et al., 2011). Hence, setting the elderly status variable to zero for nonelderly groups maximizes the possibility of observing a positive postsleep gain at the model intercept.

The linear and quadratic effect of *time of testing* is a property of performance rather than a methodological confound, so the minimally confounding value is not defined. Given our working hypothesis that there is no consolidation-based postsleep gain effect, we protected against false confirmation by setting the intercept to the time of testing that corresponds to the largest predicted postsleep gain (2 p.m.; 14:00 hr). If β_0 is nonsignificant or negative for that case, then we can infer that the postsleep gain estimate is also nonsignificant, or is negative, at any other time of testing.

In summary, the intercept for the Case 3 analyses provides an estimated true postsleep gain effect for the following conditions: sleep groups, nonelderly participants, zero data averaging, block duration of 10 s, training duration of 360 s, and 2 p.m. *time of testing*. Model intercept predictions for any other combination of the variable values can be calculated using the reported regression

coefficients. Note that the linear shifts in the values of the moderator variable intercepts that are described here have no impact on the parameter estimates for either the experimental design moderator variable effects (i.e., the α_k values) or the publication bias estimate, β_1.

Results

Comparison of HRE and WLS Analyses

Prior to investigating publication bias, we compared HRE regression results for the 88 effects sizes from the 2015 dataset (Pan & Rickard, 2015), with all previously identified moderators variables included, to WLS regression results for the same dataset with all previously identified moderators included, but without inclusion of the publication bias estimator (i.e., using Equation 3, but without the *SE* moderator). Results are shown in Table 2. The critical finding is that the moderator parameter estimates and standard errors were very similar in those two analyses. That equivalence confirms that the moderating variables account for effect size heterogeneity to the roughly the same extent in the current WLS analysis as in the prior HRE analyses.

PET Analyses for Cases 1 Through 3

We first explored publication bias using the PET Equations, for each of three cases. In Case 1, no moderating variables were included (see Equation 1). This is the most common application of PET–PEESE in the literature and it provides a reference publication bias estimate for comparison to the more critical analyses in Cases 2 and 3. The results for Case 1 suggest severe publication bias (see Table 3). That effect is visually evident in Figure 2. There is an absence of effects in the lower right area of the figure (as expected if there is publication bias), and a robust effect of standard error ($p < .0001$), as indicated in the figure by the slope of the prediction line. We next explored the case of only sleep-status as a moderator variable (Case 2 of Table 3), using PET Equation 3. The effect of standard error was again substantial. Further, in agreement with the prior meta-analysis, a sleep-status (i.e., relative sleep gain) effect was observed ($d = -.47$, $p = .001$).

Finally, we performed the PET analysis with all moderating variables included (Case 3; see Equation 3), which substantially reduces the unaccounted-for effect size heterogeneity. Results are summarized in Table 3. The standard error parameter estimate was once again robust and statistically significant ($p < .0001$). Indeed, from the smallest (.15) to the largest (.98) effect size standard error value in the dataset, the β_1 estimate of 1.95 predicts a very large increase in *d* of about 1.6; that is, $1.95 \times (.98 - .15) = 1.6$. Hence, there is strong evidence for publication bias in the dataset. In contrast, the estimated sleep-status effect is small and nonsignificant ($d = .17$, $p = .16$), suggesting that there may be a negligible relative gain effect after adjusting for both experimental design confounds and publication bias.

PEESE Intercept Estimate for Case 3

Because the intercept estimate in the PET analysis was statistically significant, the third step in the PET–PEESE analysis is application of PEESE (see Equation 4) to estimate the true effect size at the intercept after adjusting for publication bias. That analysis yielded $d = -.43$, $p = .0007$. Thus, when both the full set of previously identified moderator variables and *SE* are fitted simultaneously, the results imply a negative postsleep gain (i.e., worsened performance after the delay), rather than the positive gain that is anticipated from the primary literature.

Review of the More Recent Literature

We found 12 articles that were published after the cutoff date for inclusion in the 2015 meta-analysis, and met the experimental design inclusion criteria used by those authors. Formal meta-analysis with moderator variables was not performed for those articles, because (a) for several articles it was not possible to extract the necessary statistics for individual experimental groups that are needed for that analysis, (b) there was minimum variability in the values of the major moderator variables that we have previously identified, which, combined with the small set of articles, negated any possibility of accurately estimating their moderating effect on

Table 2
Comparison of HRE Versus Standard WLS Results (Without Standard Error as a Moderator)

		HRE			WLS		
Analysis type	Moderator/intercept	Effect size estimate (*d*)	*SE*	*p*	Effect size estimate (*d*)	*SE*	*p*
Case 1: No moderators	Intercept	0.83	0.10	<.0001	0.77	0.079	<.0001
Case 2: Sleep status moderators included	Intercept	1.00	0.12	<.0001	0.93	0.087	<.0001
	Sleep status	−0.62	0.17	.0022	−0.56	0.17	.0012
Case 3: All moderators included	Intercept	−0.29	0.19	.19	−0.31	0.18	.087
	Sleep status	−0.26	0.080	.012	−0.26	0.14	.073
	Data averaging	0.013	0.0021	<.0001	0.013	0.0,021	<.0001
	Performance duration	0.032	0.011	.049	0.034	0.010	.0011
	Training duration	−0.0014	0.0005	.031	−0.0,014	0.0005	.0071
	Time of testing	0.0036	0.015	.82	0.0032	0.014	.82
	Time of testing, squared	−0.014	0.0032	<.0001	−0.014	0.0025	<.0001
	Elderly status	−1.61	0.20	.0034	−1.64	0.21	<.0001

Note. Analyses were performed on the full dataset of the 88 group, including 65 sleep and 23 wake groups. HRE = hierarchical random effects; WLS = weighted ordinary least squares.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 3
PET Results for the Full Dataset

Analysis type	Moderator/intercept	Effect size estimate (*d*)	*SE*	*p*
Case 1: No moderators	Intercept	−0.28	0.19	.15
	SE	3.05	0.59	<.0001
Case 2: Sleep status moderator included	Intercept	−0.19	0.18	.31
	SE	3.16	0.56	<.0001
	Sleep-status	−0.47	0.14	.0,010
Case 3: All moderators included	Intercept	−0.70	0.14	<.0001
	SE	1.94	0.42	<.0001
	Sleep-status	−0.17	0.12	.16
	Data averaging[a]	0.0098	0.0018	<.0001
	Performance duration[a]	0.030	0.0071	<.0001
	Training duration[a]	−0.00093	0.00042	.032
	Time of testing[b]	−0.011	0.013	.39
	Time of testing, squared[b]	−0.014	0.0020	<.0001
	Elderly status	−1.41	0.17	<.0001

[a] In seconds. [b] In hours.

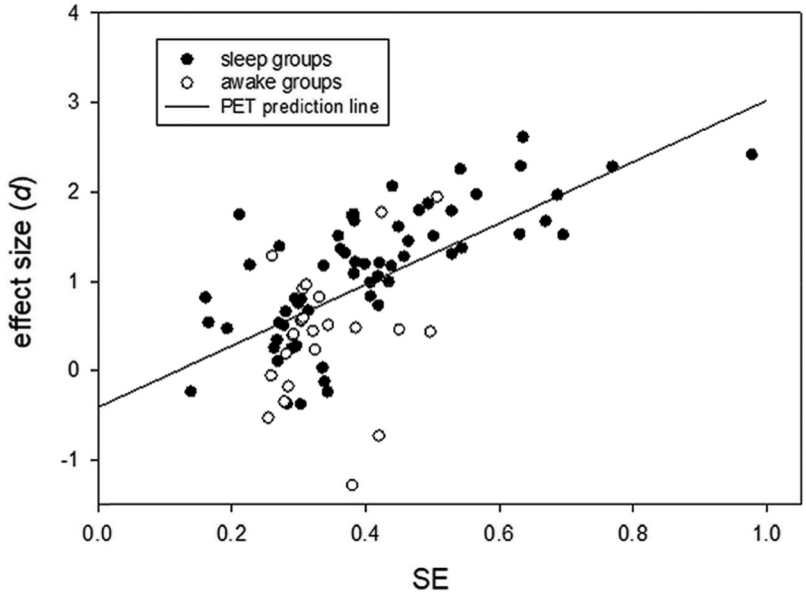
performance, and (c) a formal meta-analysis to simply estimate the aggregate effect size for the postsleep and relative gains would include only a subset of the articles (see a the preceding text), we deemed that approach less useful than the alternative approach described below, which was applicable to all reported effects from all 12 articles.

For each article, the authors' conclusions regarding both postsleep gain and relative sleep gain (where applicable) were coded on a four-level scale which captured both the direction of the trend and statistical significance (see Table 4). For the postsleep gain, *positive* refers to better performance at the beginning of the test than at the end of training. For the relative sleep gain, *positive* refers to the finding of a gain score that is larger for the sleep group than for the wake group.

Results for postsleep gain are generally in-line with the earlier literature. A positive gain effect was observed in 18 of the 20 statistical tests, 14 of which were statistically significant. For the relative gain, 10 of 15 results were positive, six of them significantly so. In summary, the postsleep and relative gain results for the newer studies are similar to those of the older studies.

Critical to interpretation of those results is whether the experimental design and data analysis that have produced illusory sleep gain effects in the literature have been improved in the new studies. Unfortunately, in nearly all cases they were not. First, group sample sizes remain small (*Mdn* = 14 participants), perpetuating the fertile conditions for publication bias as demonstrated in the current article. Second, the previously identified experimental design confounds remain. In fact, in all studies listed in Table 4, the experimental design and data analysis were identical to, or nearly identical to, those that dominated in the earlier studies. Third, data analysis has not changed in most of the articles. As in

Figure 2
Relationship Between Standard Error and Effect Size Among 88 Effect Sizes With No Moderating Variables Included in the Analysis



This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 4
Recent Studies of Sleep and Explicit Motor Learning

Source	*M* group size	Group	Postsleep gain	Relative gain
Astill et al. (2014)	30	Children, 12-hr sleep vs. wake	Positive	Positive
		Children, 24-hr sleep group	Positive	—
Gregory et al. (2014)	10	Adults, 12-hr sleep vs. wake	Positive	Positive
		Adults, 24-hr sleep (with fMRI)	Positive	—
Gudberg et al. (2015)	11.4	Young adults, sleep vs. wake	Positive	Positive
		Older adults, sleep vs. wake	Positive[a]	Positive[a]
Backhaus et al. (2016)	11.7	Adults, nap vs. wake	Positive[a]	Negative[a]
		Adults, long nap vs. wake	Positive[a]	Negative[a]
Backhaus et al. (2016)	11	Older adults, nap vs. wake	Negative[a]	Negative[a]
		Older adults, long nap vs. wake	Negative[a]	Negative[a]
Bottary et al. (2016)	18.3	Young adults, sleep vs. wake	Positive	Positive
		Older adults, sleep vs. wake	Positive[a]	Positive[a]
Cedernaes et al. (2016)	16	Adults, 8.5-hr sleep ("NSS")	Positive	—
		Adults, 4.25-hr sleep ("SSS")	Positive	—
Tucker et al. (2016)	10	Sleep vs. wake	Positive	Positive
Vien et al. (2016)	14.3	Young, nap vs. No nap	Positive	Positive
		Old, nap vs. No nap	Positive	Positive
Wamsley et al. (2016)	24.3	Sleep (combined) vs. wake (combined)	Positive	Positive[a]
Fogel et al. (2017)	13	Sleep	Positive	—
Maier et al. (2017)	18	Sleep (combined) vs. wake (combined)	Positive	Positive[a]

Note. All reported gains are significant at $\alpha = .05$, unless otherwise noted. A dash indicates that the data were not obtained.
[a] Nonsignificant result.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

the earlier studies, it involves averaging of results across the last few training blocks to compute a pretest result, and over the first few test blocks to compute a posttest result, constituting a potential online learning confound that has been confirmed in our meta-analyses (see also Pan & Rickard, 2015). In summary, all the previously identified confounding factors remain in the more recent literature. Thus, the results for those studies do not contradict our conclusions based on the earlier studies.

One article listed in Table 4 (Maier et al., 2017) compared performance across wake and sleep groups in a design involving 12 training and 12 test blocks. They did not observe a statistically significant relative sleep gain effect in the usual the comparison of the last few training blocks to the first few test blocks (as indicated in Table 4), but they did observe a relative gain in the comparison of the last few training blocks to the last few test blocks, raising the intriguing possibility that sleep dependent consolidation effects may not manifest immediately, but rather only after sufficient practice. However, in other articles in which a large number of test blocks were administered (Cai & Rickard, 2009; Landry et al., 2016), there has been no pattern of increasing relative sleep gain over blocks.

Finally, there is a pattern of scholarly omission among the articles in Table 4. Our first article raising the specter of serious confounding factors in this literature was published in 2008; and two other articles making related points were published prior to 2011 (Brawn et al., 2010; Sheth et al., 2008). In contrast, all the articles listed in Table 4 were published during or after 2014. Despite that time differential, none of that prior work was discussed among the nine articles listed in Table 4 in which the authors concluded in favor of sleep consolidation without reservation. That pattern of noncitation is remarkable considering the virtual absence of direct challenges in the literature to the evidence for serious confounds, the sole exception being a critique of the 2015 meta-analysis (Adi-Japha & Karni, 2016), to which we responded (Rickard & Pan, 2017). In contrast, in each of the three articles in Table 3 in which the authors did *not* observe an expected gain effect, some of those prior articles were cited and discussed (Backhaus et al., 2016; Maier et al., 2017). It seems unlikely that such a strong relationship between the authors' conclusions and whether the prior work was cited would occur by chance.

Examples of Studies That Used Improved Experimental Designs

Several studies in the literature have used improved experimental designs that mitigate the effects of one or more confounding factors. In those studies, results were consistent with our current conclusions. Here we summarize three of those studies. Rickard et al. (2008; see that Experiment 2) used a spaced training design with the goal of reducing both block-level reactive inhibition and session-wide fatigue. In their spaced training group, 36 training blocks with 10 s duration were interleaved with 30 s rest periods, yielding 6 minutes of time on task. Their "massed training" control group had the typical design in the literature: 12 training blocks with 30-s duration, interleaved with 30-s rest periods, again yielding 6 min of time of task. To avoid time of day confounds, both groups were trained at 1 p.m. and tested 24 hr later. Both groups underwent normal sleep between the training and test sessions. Rickard et al., avoided the inherent data averaging confound by using curve fitting and a continuity test to estimate the postsleep gain. They observed the usual postsleep gain in the massed training group. However, there was no postsleep gain in a spaced training group.

Brawn et al. (2010) also compared massed training with spaced training, using a 12-hr delay between training and the posttest sessions. Unlike the Rickard et al. (2008) study, they also included a 5 min rest period immediately after the end of the training phase, followed by a postrest test involving two blocks. Their results for the spaced training group replicated those of Rickard et al., indicating no postsleep gain effect. For the massed training group, Brawn et al., demonstrated substantially improved performance between the end of training and the postrest test and a postsleep gain between

the end of training and the final posttest. Critically, however, there was no evidence for a postsleep gain in their comparison between the postrest test and the final posttest.

The combined results of Rickard et al. (2008) and Brawn et al. (2010) demonstrate that either spaced training or a rest period after training can achieve the same result in different ways. Spaced training appears to prevent the confounding build-up of reactive inhibition and session wide fatigue, whereas a posttraining rest period resolves the build-up of those effects that occur during massed training. Those two manipulations thus constitute converging evidence against the claim of a sleep dependent consolidation process that enhances performance.

Landry et al. (2016) used the typical massed training design, a 10 min post training rest followed by a postrest test (as in Brawn et al., 2010), and a final posttest seven hours later. Their sleep groups involved a 2 hr nap rather than a full night of sleep. As in Brawn et al. (2010), they observed marked performance improvement between the post training test and the postrest test, but no improvement between the postrest test and the final posttest. They observed that pattern in each of four delay groups: uninterrupted nap, fragmented nap, quiet wake, and active wake. Thus, the Landry et al., results yielded no evidence of either a postsleep gain or a relative gain.

Discussion

In prior work, confounding design and analysis variables in the sleep and motor learning literature were identified. In the current work, we have established publication bias as an additional factor that complicates interpretation. When we simultaneously accounted for both types of confounds in the current meta-analysis, we found that the literature suggests no, or possibly a negative, postsleep gain, and at best a small relative gain.

The estimated sleep-status (relative gain) effect in the Case 3 PET analyses is smaller than in our prior HRE analysis, and it was no longer statistically significant. That result raises the possibility that one type of publication bias in this literature is unpublished experiments involving matched sleep and wake group pairs in which the relative gain effect was not statically significant. As a consequence, the average relative gain effect in the published literature would be inflated. That matched wake-sleep group publication bias would be expected to yield both inflated postsleep gains effects and deflated postwake gains in the published literature. Publication bias is likely also present among studies involving only sleep groups.

We believe that the current and prior work makes a compelling case against a postsleep performance gain, and hence against sleep-dependent consolidation that enhances motor skill in an absolute sense. The results do not rule out a small relative gain effect, and hence they do not rule out sleep-dependent consolidation in the form of protection from forgetting. In the context of apparent extensive publication bias in this literature, however, that conclusion is not yet compelling in our view.

The Perniciousness of Publication Bias and Recommendations

Between the two issues summarized above, that of publication bias may be the more problematic. The identified design and analysis confounds do not call into question the replicability of published results. If those factors constituted the only sources of bias, then new studies involving improved design and analysis would allow for convergence on the true effects. The misleading influence of publication bias, on the other hand, may be difficult to eliminate unless the great majority of future, well-designed studies are published regardless of results, a goal that would require broad cooperation among both researchers and publication outlets.

Multiple articles provide general recommendations for the reduction of publication bias (Asendorpf et al., 2013; Button et al., 2013; Simmons et al., 2011). In the sleep and motor learning literature, a sample size of at least 40 participants per group would be prudent, although statistical power analysis to estimate required sample size is preferred and should be required by journals. It should not be difficult to access existing data sets to obtain variance estimates that will support a priori power analysis. Further, the apparent absence of sleep effects among studies that are relatively well-controlled for confounding factors motivates power analysis for a two-tailed test. Finally, authors should publish both null and contradictory results (Engel & Matosin, 2014; Franco et al., 2014). If such findings cannot be published in a primary journal, they should be published in one of the several outlets that expressly embrace publication of null results.

Finally, our results raise the possibility that publication bias and experimental design flaws are problematic in other memory consolidation literatures, including those that address other motor tasks, declarative memory tasks, and animal studies. It would behoove researchers to consider that possibility sooner rather than later, particularly if small samples predominate.

Limitations of the Current Work

As with any application of meta-analysis with moderating variables, there may be patterns in the data that were not accounted for. Although inclusion of the full set of moderating variables accounted for the majority of the heterogeneity among the 88 effects sizes, a portion of the between-article heterogeneity remains unexplained. However, our conclusion that substantial publication bias exists in this literature is, in our view, unlikely to be compromised by that fact. The publication bias effect (β_1) was robust, both when the PET analysis included no moderating variables (and hence none of the effect size heterogeneity was accounted for) and when it included the full set of moderators (and hence most of the effect size heterogeneity was accounted for). In both cases, the p value for β_1 was less than .0001. The fact that the β_1 estimate remained robust when most of the effect size heterogeneity was accounted for suggests that it would also remain robust in the hypothetical case in which the remaining between article heterogeneity is accounted for.

With respect to the PEESE estimate for β_0 in the Case 3 analysis, which suggests forgetting over the delay between sessions, we should be more cautious. On one hand, the prospect of forgetting even after a sleep period is plausible, because it is not possible to eliminate some confounding influences, particularly a build-up of general fatigue and reactive inhibition during training (i.e., it is not possible to assuredly reduce either confounding effect to zero, because there must be training to produce task learning). On the other hand, simulation results (Alinaghi & Reed, 2018) show that residual, unaccounted for between-article heterogeneity can inflate the magnitude of the β_0 estimate. Currently, the strongest

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

inference regarding the true postsleep gain effect comes from a combination of the meta-analytical results and the results of experiments in the literature in which efforts have been made to reduce confounding influences (Brawn et al., 2010; Landry et al., 2016; Nettersheim et al., 2015; Rickard et al., 2008). Those experiments have yielded nonsignificant postsleep-gain estimates that exhibit no positive trend.

We have focused here on behavioral results for humans. Our results do not speak directly to the possibility of sleep-based motor skill consolidation in other animals. The current results also do not address electrophysiological or neuroimaging results for humans. However, the neurophysiological evidence for sleep consolidation in the motor sequence task is mixed. In our 2015 meta-analysis (Pan & Rickard, 2015), the EEG results of eight articles that were included in behavioral meta-analysis were reviewed. In those articles, correlations between various types of EEG signals and the magnitude of sleep-gain effects were calculated. Null results predominated, despite the broad use of uncorrected multiple comparisons. In more recent neurophysiological and brain imaging studies (e.g., Studte et al., 2017), correlations have been demonstrated between brain activations and behavioral sleep effects. However, most of those studies have employed the same problematic experimental designs as have the majority of behavioral studies. Various identified confounding factors that can differentially affect performance during training versus a test phase, or for wake versus sleep groups, can also potentially account for the neurophysiological effects. Further, neurophysiological studies nearly always involve small sample sizes, raising the prospect of publication bias in that subliterature. Because there is apparently no behavioral postsleep gain—and equivocal relative gain—among humans when confounding factors that are unrelated to consolidation processes are corrected, any argument that neurophysiological findings reflect memory consolidation is currently unpersuasive.

Conclusions

We have presented evidence of severe publication bias in the sleep and motor sequence learning literature. Using PET–PEESE analyses with moderator variables that account for effect size heterogeneity, we found no evidence for a postsleep gain and inconclusive evidence for a relative gain. Going forward, improved experimental methods, along with efforts to increase sample size and report null or contradictory results, will be needed to rigorously test the hypothesis of a sleep-dependent consolidation process in human motor sequence learning.

References

Adi-Japha, E., & Karni, A. (2016). Time for considering constraints on procedural memory consolidation processes: Comment on Pan and Rickard (2015) with specific reference to developmental changes. *Psychological Bulletin*, *142*(5), 568–571. https://doi.org/10.1037/bul0000048

Alinaghi, N., & Reed, W. R. (2018). Meta-analysis and publication bias: How well does the FAT-PET–PEESE procedure work? *Research Synthesis Methods*, *9*(2), 285–311. https://doi.org/10.1002/jrsm.1298

Asendorpf, J. B., Conner, M., Fruyt, F. D. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Astill, R. G., Piantoni, G., Raymann, R. J. E. M., Vis, J. C., Coppens, J. E., Walker, M. P., Stickgold, R., Van Der Werf, Y. D., & Van Someren, E. J. W. (2014). Sleep spindle and slow wave frequency reflect motor skill performance in primary school-age children. *Frontiers in Human Neuroscience*, *8*, 910. https://doi.org/10.3389/fnhum.2014.00910

Backhaus, W., Braaß, H., Renné, T., Krüger, C., Gerloff, C., & Hummel, F. C. (2016). Daytime sleep has no effect on the time course of motor sequence and visuomotor adaptation learning. *Neurobiology of Learning and Memory*, *131*, 147–154. https://doi.org/10.1016/j.nlm.2016.03.017

Backhaus, W., Kempe, S., & Hummel, F. C. (2016). The effect of sleep on motor learning in the aging and stroke population: A systematic review. *Restorative Neurology and Neuroscience*, *34*(1), 153–164. https://doi.org/10.3233/RNN-150521

Borragán, G., Urbain, C., Schmitz, R., Mary, A., & Peigneux, P. (2015). Sleep and memory consolidation: Motor performance and proactive interference effects in sequence learning. *Brain and Cognition*, *95*, 54–61. https://doi.org/10.1016/j.bandc.2015.01.011

Bottary, R., Sonni, A., Wright, D., & Spencer, R. M. C. (2016). Insufficient chunk concatenation may underlie changes in sleep-dependent consolidation of motor sequence learning in older adults. *Learning & Memory*, *23*(9), 455–459. https://doi.org/10.1101/lm.043042.116

Brawn, T. P., Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2010). Consolidating the effects of waking and sleep on motor-sequence learning. *The Journal of Neuroscience*, *30*(42), 13977–13982. https://doi.org/10.1523/JNEUROSCI.3295-10.2010

Breton, J., & Robertson, E. M. (2017). Dual enhancement mechanisms for overnight motor memory consolidation. *Nature Human Behaviour*, *1*(6), 0111. https://doi.org/10.1038/s41562-017-0111

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. https://doi.org/10.1038/nrn3475

Cai, D. J., & Rickard, T. C. (2009). Reconsidering the role of sleep for motor memory. *Behavioral Neuroscience*, *123*(6), 1153–1157. https://doi.org/10.1037/a0017672

Cedernaes, J., Sand, F., Liethof, L., Lampola, L., Hassanzadeh, S., Axelsson, E. K., Yeganeh, A., Ros, O., Broman, J.-E., Schiöth, H. B., & Benedict, C. (2016). Learning and sleep-dependent consolidation of spatial and procedural memories are unaltered in young men under a fixed short sleep schedule. *Neurobiology of Learning and Memory*, *131*, 87–94. https://doi.org/10.1016/j.nlm.2016.03.012

Diekelmann, S. (2017). Neuroscience: Sleep, memories, and the brain. *Nature Human Behaviour*, *1*(6), 1–2. https://doi.org/10.1038/s41562-017-0124

Engel, M., & Matosin, N. (2014, September 24). Positives in negative results: When finding "nothing" means something. *The Conversation.* http://theconversation.com/positives-in-negative-results-when-finding-nothing-means-something-26400

Fogel, S., Albouy, G., King, B. R., Lungu, O., Vien, C., Bore, A., Pinsard, B., Benali, H., Carrier, J., & Doyon, J. (2017). Reactivation or transformation? Motor memory consolidation associated with cerebral activation time-locked to sleep spindles. *PLoS ONE*, *12*(4), e0174755. https://doi.org/10.1371/journal.pone.0174755

Franco, A., Malhotra, N., & Simonovits, G. (2014). Social science. Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502–1505. https://doi.org/10.1126/science.1255484

Gregory, M. D., Agam, Y., Selvadurai, C., Nagy, A., Vangel, M., Tucker, M., Robertson, E. M., Stickgold, R., & Manoach, D. S. (2014). Resting state connectivity immediately following learning correlates with subsequent sleep-dependent enhancement of motor task performance. *NeuroImage*, *102*(2), 666–673. https://doi.org/10.1016/j.neuroimage.2014.08.044

Gudberg, C., Wulff, K., & Johansen-Berg, H. (2015). Sleep-dependent motor memory consolidation in older adults depends on task demands. *Neurobiology of Aging*, *36*(3), 1409–1416. https://doi.org/10.1016/j.neurobiolaging.2014.12.014

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Humiston, G. B., & Wamsley, E. J. (2018). A brief period of eyes-closed rest enhances motor skill consolidation. *Neurobiology of Learning and Memory*, *155*, 1–6. https://doi.org/10.1016/j.nlm.2018.06.002

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124

King, B. R., Saucier, P., Albouy, G., Fogel, S. M., Rumpf, J.-J., Klann, J., Buccino, G., Binkofski, F., Classen, J., Karni, A., & Doyon, J. (2017). Cerebral activation during initial motor learning forecasts subsequent sleep-facilitated memory consolidation in older adults. *Cerebral Cortex*, *27*(2), 1588–1601. https://doi.org/10.1093/cercor/bhv347

Korman, M., Raz, N., Flash, T., & Karni, A. (2003). Multiple shifts in the representation of a motor sequence during the acquisition of skilled performance. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 12492–12497. https://doi.org/10.1073/pnas.2035019100

Landry, S., Anderson, C., & Conduit, R. (2016). The effects of sleep, wake activity and time-on-task on offline motor sequence learning. *Neurobiology of Learning and Memory*, *127*, 56–63. https://doi.org/10.1016/j.nlm.2015.11.009

Maier, J. G., Piosczyk, H., Holz, J., Landmann, N., Deschler, C., Frase, L., Kuhn, M., Klöppel, S., Spiegelhalder, K., Sterr, A., Riemann, D., Feige, B., Voderholzer, U., & Nissen, C. (2017). Brief periods of NREM sleep do not promote early offline gains but subsequent on-task performance in motor skill learning. *Neurobiology of Learning and Memory*, *145*, 18–27. https://doi.org/10.1016/j.nlm.2017.08.006

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V., & Howard, J. H., Jr. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research*, *201*(2), 351–358. https://doi.org/10.1007/s00221-009-2024-x

Nettersheim, A., Hallschmid, M., Born, J., & Diekelmann, S. (2015). The role of sleep in motor sequence consolidation: Stabilization rather than enhancement. *The Journal of Neuroscience*, *35*(17), 6696–6702. https://doi.org/10.1523/JNEUROSCI.1236-14.2015

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pan, S. C., & Rickard, T. C. (2015). Sleep and motor learning: Is there room for consolidation? *Psychological Bulletin*, *141*(4), 812–834. https://doi.org/10.1037/bul0000009

Rickard, T. C., Cai, D. J., Rieth, C. A., Jones, J., & Ard, M. C. (2008). Sleep does not enhance motor sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 834–842. https://doi.org/10.1037/0278-7393.34.4.834

Rickard, T. C., & Pan, S. C. (2017). Time for considering the possibility that sleep plays no unique role in motor memory consolidation: Reply to Adi-Japha and Karni (2016). *Psychological Bulletin*, *143*(4), 454–458. https://doi.org/10.1037/bul0000094

Sheth, B. R., Janvelyan, D., & Khan, M. (2008). Practice makes imperfect: Restorative effects of sleep on motor learning. *PLoS ONE*, *3*(9), e3190. https://doi.org/10.1371/journal.pone.0003190

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simor, P., Zavecz, Z., Horváth, K., Éltető, N., Török, C., Pesthy, O., Gombos, F., Janacsek, K., & Nemeth, D. (2018). Deconstructing procedural memory: Different learning trajectories and consolidation of sequence and statistical learning. *Frontiers in Psychology*, *9*, 2708. https://doi.org/10.3389/fpsyg.2018.02708

Stanley, T. D. (2017). Limitations of PET–PEESE and other meta-analysis methods. *Social Psychological & Personality Science*, *8*, 581–591. https://doi.org/10.1177/1948550617693062

Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods*, *5*(1), 60–78. https://doi.org/10.1002/jrsm.1095

Studte, S., Bridger, E., & Mecklinger, A. (2017). Sleep spindles during a nap correlate with post sleep memory performance for highly rewarded word-pairs. *Brain and Language*, *167*, 28–35. https://doi.org/10.1016/j.bandl.2016.03.003

Tucker, M., McKinley, S., & Stickgold, R. (2011). Sleep optimizes motor skill in older adults. *Journal of the American Geriatrics Society*, *59*(4), 603–609. https://doi.org/10.1111/j.1532-5415.2011.03324.x

Tucker, M. A., Nguyen, N., & Stickgold, R. (2016). Experience playing a musical instrument and overnight sleep enhance performance on a sequential typing task. *PLoS ONE*, *11*(7), e0159608. https://doi.org/10.1371/journal.pone.0159608

Vien, C., Boré, A., Lungu, O., Benali, H., Carrier, J., Fogel, S., & Doyon, J. (2016). Age-related white-matter correlates of motor sequence learning and consolidation. *Neurobiology of Aging*, *48*, 13–22. https://doi.org/10.1016/j.neurobiolaging.2016.08.006

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: Sleep-dependent motor skill learning. *Neuron*, *35*(1), 205–211. https://doi.org/10.1016/S0896-6273(02)00746-8

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. https://doi.org/10.1101/lm.58503

Wamsley, E. J., Hamilton, K., Graveline, Y., Manceor, S., & Parr, E. (2016). Test Expectation Enhances Memory Consolidation across Both Sleep and Wake. *PLoS ONE*, *11*(10), e0165141. https://doi.org/10.1371/journal.pone.0165141

Received April 8, 2021
Revision received August 5, 2021
Accepted August 21, 2021 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.