Sleep Consolidation of Musical Competence

Stephen C. Van Hedger
University of Chicago

Anders Hogstrom
University of Chicago

Caroline Palmer
McGill University, Montreal, Canada

Howard C. Nusbaum
University of Chicago

Previous research has shown that sleep facilitates skill acquisition by consolidating recent memories into a stabilized form. The way in which sleep benefits the acquisition of a musical skill, however, is unclear. This is because previous studies have not dissociated the extent to which sleep consolidates learned motoric patterns from the conceptual structure of the music. We thus designed two experiments in which pianists performed short pieces – designed to separate conceptual from motoric errors – over the course of a day. In Experiment 1, participants were trained in the morning and tested immediately, 12 hours, and 24 hours after training. While both motor and conceptual errors increased over a waking retention interval, only conceptual errors were significantly reduced after sleep. Moreover, individuals who reported spending more time "playing by ear" showed greater reductions of conceptual errors after sleep. A second control experiment, in which participants were trained in the evening and tested immediately – as well as 12 hours – after training, confirmed that the results from Experiment 1 could not be attributed solely to circadian confounds or to elapsed time since training. Together, these results suggest that conceptual and motor errors consolidate differently and interact with differences in practice style.

Skill acquisition involves a consolidation process, whereby newly formed memories of repeated experiences are consolidated into a more stabilized form (McGaugh, 2000). This consolidation process, moreover, has been shown to benefit from sleep (e.g., Diekelmann, Wilhelm, & Born, 2009; Walker & Stickgold, 2006). Benefits of sleep consolidation have been studied in multiple contexts, including perceptual learning of synthetic speech (Fenn, Nusbaum, & Margoliash, 2003), auditory discrimination (Atienza & Cantero, 2001), and learning motor sequences (e.g., Brawn, Fenn, Nusbaum, & Margoliash, 2010; Walker, Brakefield, Hobson, & Stickgold, 2003).

Learning a skill, however, is not simply a matter of learning specific rote motor patterns or recognizing particular sensory patterns. The abstraction of rules is also central, as it supports generalization beyond specific experiences. Rule abstraction has also been tied to sleep-dependent consolidation. In statistical learning paradigms, Durrant, Taylor, Cairney, and Lewis (2011) found a facilitation of rule learning following sleep, compared to a waking retention interval of the same duration. After sleep, participants were significantly better at determining whether novel tone sequences followed statistical rules that were presented before sleep, compared to a control group. Furthermore, the abstraction of statistical regularities – which has been promoted as a mechanism of language acquisition – has been shown to benefit infants through naps (Gómez, Bootzin, & Nadel, 2006).

In Western tonal music, pitch hierarchies and harmonic relationships also follow statistical regularities (e.g., Huron, 2006; Jonaitis & Saffran, 2009; Temperley, 2007). In learning to perform a new piece of music, moreover, statistical knowledge of tonal and harmonic relationships among pitches influences the nature of performance errors. Without this knowledge, pitch errors (playing a different pitch than is intended) would be unaffected by the context of a musical piece, except to the extent that the sequence permits physical (hand and finger) movement errors due to action similarity. However, pianists make more pitch errors that are congruent with the piece's key signature and harmonic structure than expected by chance (Palmer & van de Sande, 1993,

1995). By analyzing the types of errors musicians make as they learn new pieces of music and the subsequent retention of different types of errors over different periods, we can investigate how sleep consolidation interacts with different aspects of pattern knowledge underlying Western music.

A small number of studies have investigated how sleep can affect learning new pieces of music. However, these studies generally used very simple one-handed keyboard melodies as stimuli (e.g., Allen, 2012; Duke & Davis, 2006; Simmons & Duke, 2006), making performance akin to musical finger tapping. In fact, note sequences used by Duke and Davis (2006) were identical to finger-tapping sequences used by Walker et al. (2003), and consisted of the first four notes of a whole-tone (i.e., evenly spaced) scale, which does not have the same pitch hierarchies and expectations as are found in major and minor scales (cf. Krumhansl, 1979). Thus, while these sequences were performed on a piano, they did not fully capture statistical pitch regularities common in Western music. To the extent that there are such statistical regularities in the melodies of Western music, they can be thought of as providing conceptual information. As with other kinds of concepts, these concepts can guide expectations and behavior.

There is also little work that explicitly examines sleep consolidation of concept formation in motor tasks. Early studies of sleep consolidation focused on relatively simple pattern learning in order to restrict questions of consolidation to patterns that showed very little generalization (e.g., see Karni, Tanne, Rubenstein, Askensy, & Sagi, 1994). However, demonstrations that generalization learning shows consolidation (e.g., Brawn, Fenn, Nusbaum, & Margoliash, 2008; Fenn et al., 2003) suggest that the formation of more abstract memories—that are beyond the specific representation of the specific training stimuli—may be part of the consolidation process. No study has yet examined whether sleep acts to consolidate both motor learning and conceptual learning inherent in such sequences. Learning to play a piece of music requires a performer to learn both the sequence of motor actions as well as the conceptual structure of the piece.

The present study was designed to examine how sleep consolidation separately affects the motor learning and the conceptual or abstract learning for conceptually related pieces of music. The stimuli used in the present study were two-part musical pieces for piano, previously reported in a study of performance errors (Palmer, Mathias, & Anderson, 2012) that follow Western tonal conceptual structure, and require sequences of motor movements by each hand that are complex enough to be difficult to perform at fast performance rates. Thus,

these musical stimuli can be viewed as containing both conceptual and motor sequences that participants must learn in order to perform them correctly. The pieces were developed to make it possible to separate pitch errors due to motor performance (hitting the wrong key through a slip of the hand) vs. errors due to conceptualization of the melody (hitting a wrong note that is consistent with a typical melodic structure found in the pieces). In Experiment 1, pianists learned four short piano pieces in the morning, and then were tested on the same piano pieces immediately after training (in the morning), 12 hours after training (in the evening), and 24 hours later (the following morning). In the control experiment, a different group of pianists learned the same four short piano pieces in the evening, and then were tested on the same piano pieces immediately after training (in the evening) and 12 hours after training (the following morning).

If sleep consolidates both the motor and conceptual aspects of learning, and music performance follows the pattern of consolidation findings in prior research (e.g., Fenn et al., 2003), we would expect to see an increase in both kinds of errors over the course of a waking retention interval (from morning to evening in Experiment 1), with sleep acting to reduce all errors. However, Palmer and Meyer (2000) demonstrated that skilled pianists showed greater generalization across musical pieces when conceptual (melodic interval and contour) relations were retained, regardless of whether the motor movements were the same. Given these findings, it is possible that we would observe sleep consolidation for conceptual, but not specifically motor, information in piano performance. This would suggest that the nature of sleep consolidation, in the context of musical practice, has a much more specific effect on memory.

Experiment 1

METHOD

Participants. Sixteen pianists participated in the study ($M = 20.93$, $SD = 2.37$, range 18-26 years, 7 females, 9 males). Participants were specifically recruited through flyers posted in music buildings on the University of Chicago campus, as well as through e-mails sent to university piano groups. The participants had an average of 11.6 years of private instruction on the piano, and passed a prescreening measure to verify their ability to sight-read piano sheet music.

All participants were required to keep sleep logs one week prior to participating in the study (including the night of the study) to ensure consistent sleep-wake cycles. Additionally, participants reported not taking drugs or
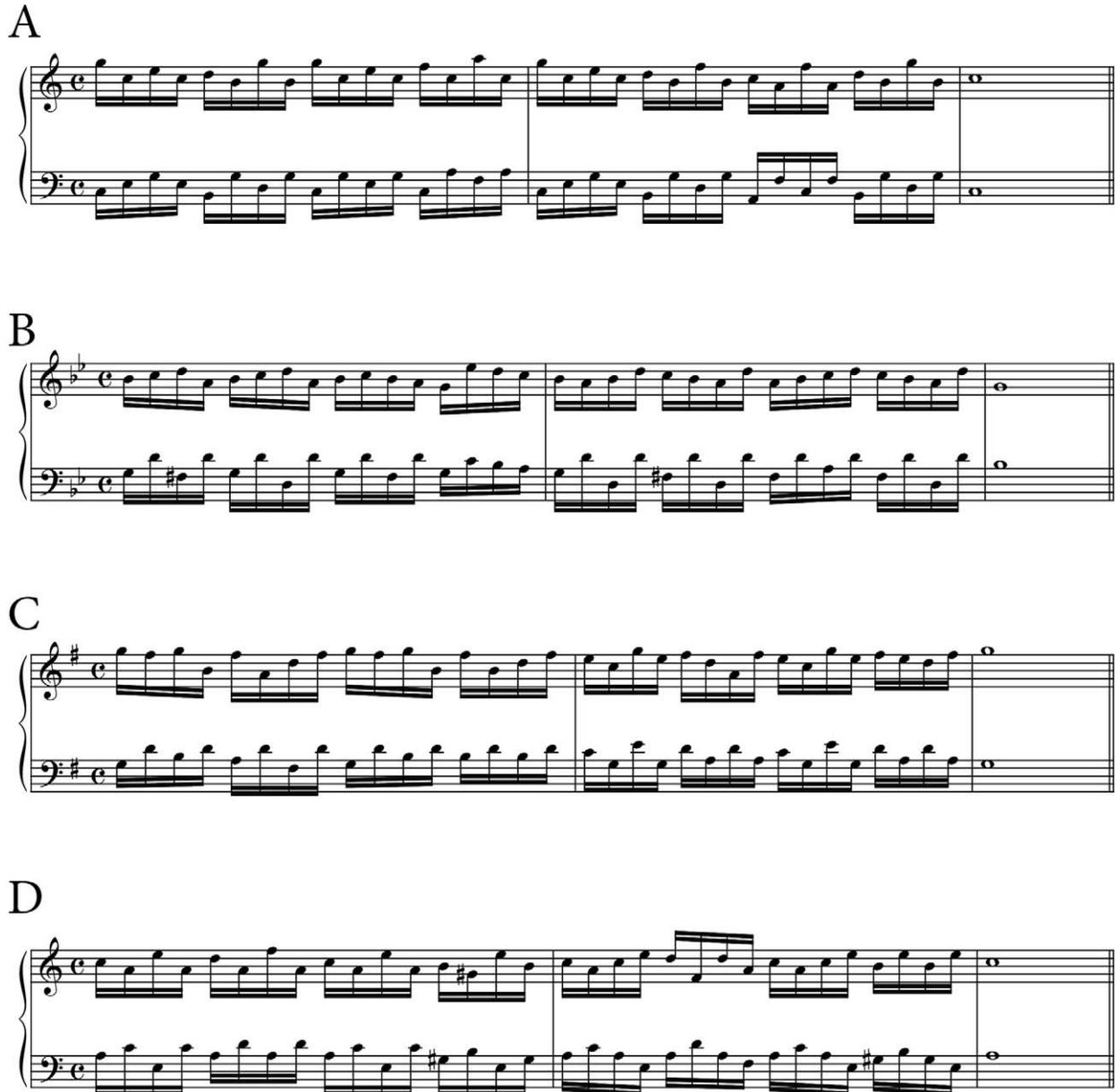
FIGURE 1. Pieces from the experiment.

substances that alter the nature or duration of sleep. This included sleeping pills, substances with a noticeable side-effect of drowsiness, and alcohol. Participants' morning routine apart from time of waking was not specifically probed (e.g., coffee consumption, breakfast, etc.).

Materials. We used four short pieces of piano music (Palmer et al., 2012), each of which consisted of continuous sixteenth notes in both the left and right hands (with the exception of the final measure, which consisted

of a whole note in both hands). Figure 1 shows all four pieces used in the experiment. Two of the pieces were in major keys (C major and G major), while the other two pieces were in minor keys (A minor and G minor).

Participants performed the pieces on a full-sized, weighted keyboard. The keyboard was connected via USB cable to a computer, which recorded MIDI information about the performances including the identity, intensity, and duration (onset and offset) of each note with temporal accuracy of 1 ms. Participants heard the

audio output corresponding to each performance through Sennheiser HD 280 headphones at a comfortable listening level (approximately 75 dB SPL).

Design. All subjects participated in all conditions. The experiment consisted of a training phase, followed by three posttests. Before each session, participants rated their sleepiness on the Stanford Sleepiness Scale (SSS), which is a 7-point subjective measure of sleepiness. Training always occurred in the morning (9am ± 1hr). The first posttest immediately followed training, while the second and third posttests occurred 12 hours and 24 hours after training, respectively.

During the training phase, participants were presented with the sheet music from the four pieces in a pseudo-randomized order (as we always followed a major key piece with a minor key piece or vice versa). Pieces were presented one at a time. Participants could spend up to five minutes learning each piece before being asked to "perform" at a slow tempo of 35 beats per minute (BPM). During these five minutes of study, participants were allowed to play each piece in an unconstrained manner (e.g., they could practice each hand separately, they could play through the piece at a fast or slow tempo, or they could focus attention on a particular phrase or section that they found to be difficult). No metronome was presented during this five minute learning session. Participants were not allowed to notate or mark the score in any way (such as writing in an accidental sign or writing in fingerings). After the five minutes had ended, or after the participant felt that they were sufficiently prepared to "perform" the piece (whichever came first), participants had to play at least three successive note-perfect repetitions to ensure that any observed differences in errors were due to experimental manipulations and not to problems in sight-reading or perceptual difficulties. On average, participants reached the performance criterion of three successive note-perfect repetitions relatively quickly ($M = 4.28$, $SD = 1.17$, range = 3-8 repetitions).

During the posttests, participants performed each of the four pieces three times at a faster tempo (55 BPM/ 60 BPM/65 BPM, in which one beat was equal to one quarter note for the immediate, 12-hour, and 24-hour posttests, respectively). Participants performed with a metronome in order to ensure they were playing at the appropriate tempo. We increased the tempo by 5 BPM after the immediate posttest to offset practice effects that might lead to speeding up, due to increased ease of performing, ensuring that participants produced measurable errors. Participants knew that the tempo would be slightly increased (by about 10%) at each posttest. While

the experimenter set the tempo for each testing session, participants always heard a one-measure metronomic lead-in before beginning the piece.

The order of the pieces was different for each posttest. For example, if a participant received the pieces in the order ABCD during training, they would play the immediate posttest as BCDA, the 12-hour posttest as CDAB, and the 24-hour posttest as DABC. Before the posttests, participants had the opportunity to warm up (e.g., play scales), but were not allowed to see the pieces, or perform any part of the pieces. Very few (2 of 16) participants took advantage of "warming up," and moreover these participants did not show any significant differences compared to other participants with regard to number of errors. At the end of the final session, participant filled out a music experience questionnaire and turned in their sleep log.

RESULTS

Given the complex nature of learning a piece of music on the piano, there are many ways in which one can operationalize performance errors (for a comprehensive account, see Palmer & van de Sande, 1993). For example, one may forget to play a written or intended note (generally called a "deletion" error), one may insert an extra note in addition to the correct note (generally called an "addition" error), or one may substitute a correct note for the incorrect note (generally called a "substitution" error). While all types of errors are informative of underlying cognitive processes, in this paper we focused on single-note substitution errors, as they most clearly reflected our underlying psychological question. Specifically, direct one-to-one substitutions can have a dissociable motor or conceptual underpinning, which is described in detail in the following paragraphs. Furthermore, in the current study, the majority of performance errors involved single notes (as opposed to chords). Substitution errors comprised 49.5% of the total number of errors ($n = 3,129$ total errors, or approximately 65 errors per subject per posttest).

Defining Motor and Conceptual Substitution Errors. We defined motor substitution errors as incorrect pitches that were ±1 semitones from the target note and therefore represented a response corresponding to a finger landing on an adjacent key, following the convention of previous research that captures slips of the hand (fingers that miss the intended spatial location; e.g., Palmer & van de Sande, 1995). In addition, motor substitution errors had to be non-diatonic, to avoid inclusion of errors that were intended (i.e., could possibly be conceptual errors because they fit the key of the musical stimulus).

The motor substitution errors were thus attributed to participants reaching for the correct key, but failing to correctly execute the intended motor plan and pressing an adjacent key instead.

Conceptual substitution errors were defined as incorrect pitches that created an error that was congruent with the key signature (called "diatonic errors" by Palmer & van de Sande, 1993) albeit incorrect in respect of the specific notation of the piece. The majority (72%) of conceptual errors were distanced by three or more semitones (i.e., several piano keys apart), providing evidence that conceptual errors were not simply motorically driven, as in a slip of the finger. Furthermore, similar to previous studies (e.g., Palmer & van de Sande, 1993), we found that the majority of pitch substitution errors (97%) resulted in diatonic outcomes. Since these errors resulted in pitches from the 7 diatonic tones of the scale on which the musical piece was based (out of a possible 12 tones), we interpret these conceptual errors as evidence of a schematization or abstraction of the pieces, since chance performance would yield a diatonic pitch error rate of 7/12, or 58%. In fact, all participants' substitution errors were above the chance rate of 58%, $X^2(1) = 16.00$, $p < .001$. In other words, participants' performance of conceptual errors reflects knowledge that the notes needed to belong to a certain scale or melodic figure, but the error conflicted with the exact order in which the notes occurred.

Within the context of each of the test pieces, conceptual and motor errors can reliably be distinguished using the definitions in the previous paragraphs. For example, if a participant incorrectly played a B instead of the intended tone C, this would be classified as a motor error (one semitone away) if the key signature of the musical piece was A-flat major (since the B outcome is *not* found in an A-flat major scale), but would be classified as a conceptual error if the key signature of the piece was C major (since B is found in a C major scale). Based on these classifications, 25% (40 of 160) of all the substitution errors that were removed from the target note by one semitone (i.e., an adjacent key) were classified as motoric. The remaining 75% of substitution errors were not considered in this analysis because they could also be due to conceptual errors given the specific note error made. While this might seem like a conservative means of defining motor substitution errors, it should be noted that expanding the definition of motor substitution errors to include *all* semitone errors (i.e., treating every semitone error as a slip of the hand) does not substantially change the present results.

Overall Substitution Errors. To test whether substitution errors showed evidence for sleep-dependent consolidation, we carried out a repeated measures analysis of variance (ANOVA). We found a significant main effect of time of posttest for overall substitution errors, $F(2, 28) = 6.84$, $p < .01$, $\eta_p^2 = .31$. Results showed performance became significantly worse over the day with participants making an average of 15.31 more errors from the immediate posttest (29.81 errors) to the 12-hour posttest after a waking retention interval (45.13 errors). However, there was a subsequent significant improvement in performance from the 12-hour posttest to the 24-hour posttest (the retention interval incorporating sleep), with participants making an average of 9.06 fewer errors in the 24-hour posttest (36.06 errors). These overall results are consistent with previous sleep consolidation results demonstrating a reduction of performance over a waking retention interval and restoration of this loss following sleep (e.g., Brawn et al., 2010). However it is possible that only one of the two types of substitution error (motor or conceptual) was driving this pattern of post-training performance change. Indeed, when including substitution error type (motor vs. conceptual) in our analysis, we observed a significant error type by posttest interaction, $F(2, 28) = 4.41$, $p = .02$, $\eta_p^2 = .24$, suggesting that the consolidation pattern of conceptual and motor errors was different. We therefore ran separate planned comparison tests to examine the effect of posttest time on the mean number of motor and conceptual errors, as this allowed us to determine whether both types of substitution errors (motor vs. conceptual) were showing evidence of sleep-dependent consolidation.

Motor Substitution Errors. Based on previous sleep consolidation research on motor sequence learning, we expected motor performance to decrease (i.e., more errors) over the waking retention interval (from immediate posttest to the 12-hour posttest), but then we expected performance to improve (i.e., fewer errors) after a night of sleep (24-hour posttest), returning to the immediate posttest error levels. For motor substitution errors, we found that participants' performance significantly worsened over the course of the day, with participants making an average of 0.94 more errors from the immediate posttest (0.69 errors) to the 12-hour posttest (1.63 errors), $t(15) = 1.86$, $p < .05$. However, we did not find evidence that sleep restored motor performance, as participants made an average of only 0.06 fewer errors from the 12-hour to 24-hour posttest (1.56 errors), which was not significantly different from the 12-hour posttest, $t(15) = 0.21$, $p > .40$, and still significantly higher than the error rate of the immediate posttest, $t(15) = 1.84$, $p < .05$.

These results are displayed in Figure 3. In other words, while we found evidence that motor substitution errors increased over a waking retention interval, sleep did not appear to actively restore performance, as the number of motor substitution errors in the 24-hour posttest (after sleep) was not statistically distinguishable from the number of motor substitution errors committed by participants after waking retention, and still significantly greater than the number of motor substitution errors committed by participants immediately after training.

In the present study, clear motor substitution errors were relatively infrequent (3% of total substitution errors) compared to the frequency of conceptual errors. While there was sufficient power to detect a reliable change in motor errors as shown by the significant error increase after waking retention, as well as significantly more motor errors the following morning compared to the immediate posttest, we repeated the analyses with an expanded definition of motor substitution errors. Specifically, we expanded the motor substitution error definition to include *all* semitone errors. While this definition of motor substitution errors could possibly also include some conceptual errors, we wanted to assess whether our pattern of motor substitution consolidation was perhaps due to an overly strict definition of motor substitution errors. Because of this expanded motor substitution error definition, we consequently contracted the conceptual substitution error definition (to include substitution errors that were two or more semitones removed from the intended target and diatonically related to the key signature), as we did not want to count the same substitution error as both a motor and conceptual substitution error.

With the expanded definition of motor substitution errors, we still observed a significant interaction between error type (conceptual and motor) and posttest, $F(2, 28) = 7.71$, $p < .01$, $\eta_p^2 = .34$, suggesting that conceptual and motor errors were not showing the same pattern of consolidation across posttests. We ran a series of planned comparison t-tests to determine whether the consolidation pattern of broadly defined motor substitution errors was similar to the consolidation pattern of strictly defined motor substitution errors. Participants made on average 2.56 broad motor substitution errors in the posttest immediately after training, but their performance significantly deteriorated over the 12-hour waking retention interval, as they made an average of 4.06 broad motor substitution errors in the 12-hour posttest, $t(15) = -2.04$, $p = .03$. Similar to the strict definition of motor substitution errors, however, broad motor errors did not seem to be helped by sleep, as participants made an average of 3.56 errors in the 24-hour posttest

incorporating sleep – a reduction that was not significantly different than the error levels of the 12-hour posttest, $t(15) = 0.57$, $p = .29$. Moreover, the broad motor substitution error rate in the 24-hour posttest was marginally higher than the broad motor substitution error rate in the immediate posttest after training, $t(15) = 1.40$, $p = .09$. Thus, even when expanding our definition of motor substitution errors to include more data points (in a manner that possibly includes some conceptually-driven errors), we observe the same general pattern of results as we did with the strictly defined motor substitution errors.

Coefficient of Variation (CV). We also examined how the coefficient of variation — CV, or standard deviation (SD) of interonset intervals (IOI) / mean IOI — for each trial varied as a function of posttest. For example, if a participant had a mean IOI of 250 ms (i.e., the average time between any two adjacent notes) and an SD (IOI) of 25 ms, their CV would be 10%, or 0.10. CV has been previously used as a general measure of motor learning (e.g., Pohl, McDowd, Filion, Richards, & Stiers, 2006), as well as a specific measure of motor precision in piano performance (e.g., MacKenzie, Vaneerd, Graham, Huron, & Willis, 1986; Repp, 1997) and finally recent work has shown that, in timing tasks, CV can be consolidated by sleep (Lewis, Couch, & Walker, 2011). Specifically, the coefficient of variation is a measure of the timing error in producing motor responses and, while different in nature from the gross execution of hitting a wrong key, if motor performance deteriorates over time, we would expect sloppier motor timing as well as key errors. It is important to note, however, that while CV can be conceptualized as a measure of motor regularity, it is possible that conceptually based errors could affect the CV (e.g., through affecting the time it takes to reach for a physically distant – but conceptually related – note on the piano). Thus, while CV certainly has a motor component, it is a complex measure that can potentially be influenced by several additional factors.

An ANOVA on CV values, with posttest as a repeated factor, revealed that participants' timing variability significantly differed across posttests, $F(2, 28) = 3.57$, $p = .04$, $\eta_p^2 = .19$. Unlike previous sleep consolidation studies using CV measurements, we did not find a significant increase in variability over the course of the waking retention interval (immediate posttest to 12-hour posttest), followed by a significant decrease of variability after sleep (12-hour posttest to 24-hour posttest). Instead, we observed no significant increase in variability over the waking retention interval [immediate posttest: $M = 7.5\%$, $SE = 0.4\%$, 12-hour posttest:
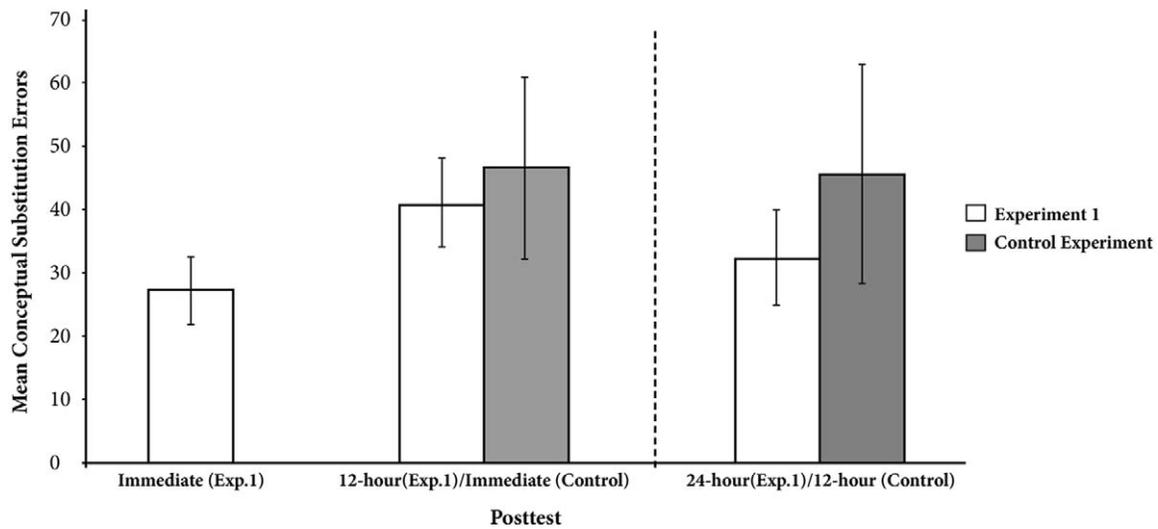
FIGURE 2. Mean conceptual substitution errors (and SE bars) as a function of posttest for both Experiment 1 and the control experiment. The dotted vertical line represents where sleep occurred.

$M = 7.6\%$, $SE = 0.5\%$; $t(15) = 0.29$, $p > .50$], and a significant *increase* in variability after sleep [24-hour posttest: $M = 8.4\%$, $SE = 0.4\%$; $t(15) = 2.67$, $p = .02$].[1]

Conceptual Errors. In contrast with the pattern of motor errors – which did not show evidence for sleep-dependent consolidation for either substitution errors or timing variability – a different pattern was observed for conceptual substitution errors. The planned comparison tests from the omnibus ANOVA revealed that similar to motor substitution errors, there was a significant increase of conceptual errors over the waking retention interval from the immediate posttest (27.25 errors) to the 12-hour posttest (41.06 errors), $t(15) = 3.43$, $p < .01$. However, unlike the motor errors, as shown in Figure 3, there was a significant reduction in

conceptual errors after a night of sleep by the 24-hour posttest (32.5 errors), $t(15) = 2.58$, $p < .05$. These results are represented in Figure 2. This is the same pattern of sleep consolidation seen in perceptual learning of synthetic speech.

Playing Styles and Consolidation. Pianists practice and perform using different strategies that could potentially affect the approach they use for encoding music into memory. Some pianists typically practice from sheet music whereas others "play by ear." Although the participants were explicitly instructed to refrain from practicing the stimulus pieces outside the experiment following initial training, these habits of playing could affect learning and performance during training in the present study. We hypothesized that practice styles that favor close attention to reading and encoding the score (e.g., primarily playing from sheet music) might show a different pattern of consolidation than styles of learning which favor using an internal mental representation (e.g., such as learning from hearing a recording, or "playing by ear"). The former might encourage a more note-by-note or phrase-based approach to learning a new piece whereas the latter might take a broader schematic approach, thereby being more at risk for more familiar schematic representations of the music (cf. Bartlett, 1932).

Participants were asked to estimate the percentage of music they typically practice by ear versus sheet music in the music experience questionnaire. We then correlated this estimated percentage of playing by ear

---

[1] To account for the increases in CV potentially due to the interaction of tempo offsets, we assessed how much the CV would change in an identical performance as a function of increasing the tempo. In theory, the CV should not change at all, as both the mean IOI and the IOI variability should change proportionally, but rounding errors might artificially increase the CV. To ensure that observed differences in CV as a function of posttest could not be simply explained by a measurement issue, we randomly selected a participant's performance and changed the tempo of the MIDI file (to 55 BPM, 60 BPM, or 65 BPM). The CV for this particular performance increased from 55 BPM (immediate posttest tempo) to the 60 BPM (12-hour posttest tempo) by 0.00067%, and increased from 60 BPM (12-hour posttest tempo) to 65 BPM (24-hour posttest tempo) by 0.011%. These increases in estimated CV were approximately 73 times smaller than the observed increase in CV, suggesting that measurement errors did not account for the observed findings. This pattern of results is shown in Figure 4.

TABLE 1. *Correlation Coefficients (p-values) of the Amount of Self-Reported Time Spent Playing By Ear and Dependent Measures*

Dependent Measure	Experiment 1	Control Experiment
Overall Substitution Errors	.02 (.94)	−.47 (.20)
Sleep Consolidation (Conceptual)	.61 (.01)*	−.10 (.80)
Sleep Consolidation (Motor)	−.01 (.97)	.25 (.52)
Sleep Consolidation (CV)	.11 (.69)	−.53 (.13)

TABLE 2. *Correlation Coefficients (p-values) of the Amount of Self-Reported Sleep and Dependent Measures*

Dependent Measure	Experiment 1	Control Experiment
Overall Substitution Errors	.19 (.48)	.16 (.68)
Sleep Consolidation (Conceptual)	−.08 (.77)	.54 (.12)
Sleep Consolidation (Motor)	−.20 (.46)	−.04 (.92)
Sleep Consolidation (CV)	−.18 (.50)	−.04 (.92)

with the magnitude of sleep-dependent consolidation. To calculate the magnitude of sleep-dependent consolidation, we subtracted the error rate of the 24-hour posttest from the error rate of the 12-hour posttest (as the retention interval between these two posttests incorporated sleep). A larger number would thus represent *more* sleep-dependent consolidation, while a smaller (or negative) number would represent *less* (or no) sleep-dependent consolidation. We ran this correlation for all three of our dependent measures (conceptual substitution errors, motor substitution errors, and CV), as well as for overall substitution errors throughout the entire experiment.

The results for all four correlations are reported in Table 1. For conceptual substitution errors, we found a significant, positive correlation between the magnitude of sleep-dependent consolidation and the reported amount of time spent playing by ear. This suggests that participants' previous strategies in learning pieces of music influenced the degree to which their conceptual substitution errors were reduced from sleep. Importantly, reported amount of time playing by ear was not significantly correlated with overall number of errors ($r = −.04$, $n = 16$, $p > .50$), years of musical experience ($r = −.08$, $n = 16$, $p > .50$), amount of practice per week ($r = .41$, $n = 16$, $p > .11$), or self-reported sight reading ability ($r = −.01$, $n = 16$, $p > .50$).

Unlike conceptual substitution errors, we did not find any evidence that reported time of playing by ear was significantly correlated with the magnitude of sleep-dependent consolidation for motor substitution errors or coefficient of variation. Finally, the reported time of playing by ear was not significantly correlated with the overall number of substitution errors.

Sleep Measures. We compared subjective sleepiness ratings (obtained from the Stanford Sleepiness Scale) for each participant across all three time points (morning training and immediate posttest, evening 12-hour posttest, and morning 24-hour posttest). This was done to assess whether participants reported significantly different sleepiness ratings as a function of time, as general differences in arousal as a function of time could potentially affect our observed results. On a scale of 1 (*feeling active and vital; alert; wide awake*) to 7 (*almost in reverie; sleep onset soon; lost struggle to remain awake*), participants gave themselves an average score of 3.16 ($SD = 1.18$) in the morning training and immediate posttest, a 2.53 ($SD = 0.76$) in the evening 12-hour posttest, and a 3.31 ($SD = 1.01$) in the morning 24-hour posttest.

A repeated-measures ANOVA revealed that these reported differences in sleepiness across session were significant, $F(2, 28) = 4.35$, $p < .05$, $\eta_p^2 = .22$, with a Least Significant Difference posthoc test showing that the evening sleepiness score was lower (i.e., participants were more awake) compared to both the first and second morning sessions ($ps = .06$ and .01, respectively). This finding is consistent with previous literature suggesting that young adults are at the acrophase of their circadian cycle during the evening (e.g., Fenn et al., 2003). It is notable, however, that young adults have been shown to perform *better* on cognitive tasks in the evening compared to the morning (e.g., Johnson et al., 1992), which is the opposite of our observed effect. Thus, it appears that we found evidence for sleep dependent consolidation in spite of the fact that our participants were more alert during the evening testing session when they showed poorer performance.

In addition to the SSS ratings, we obtained self-report sleep logs from participants, which included the night of sleep between the 12-hour and 24-hour posttests. On average, participants reported sleeping an average of 6.8 hours ($SD = 1.3$) the night of the experiment. The duration of sleep, however, was not significantly related to overall substitution errors, nor the consolidation (operationalized as the reduction of errors from the 12-hour to the 24-hour posttest) for conceptual substitution errors, the consolidation for motor substitution errors, or the consolidation for CV. These results are reported in Table 2.

Control Experiment

To rule out any circadian or arousal effects that could potentially affect the results from the first experiment,

we designed a circadian control experiment, using the same testing musical pieces as stimuli, the same general testing procedure, and different (naïve) pianists. Pianists first performed the musical pieces in the evening and we then tested their retention immediately after training as well as 12 hours later (the following morning). This change in time of testing allowed us to directly test two important hypotheses. First, if there are no major circadian effects in the first experiment, participants trained in the evening should exhibit the same patterns of motor and conceptual errors as pianists trained in the morning in Experiment 1. This would rule out the possibility that pianists simply are worse in the evenings, and cannot perform at the same proficiency as compared to the morning. While this possibility seems unlikely, the significantly different sleepiness scores in the evening posttest from Experiment 1 could suggest that arousal differences between morning and evening could affect performance (cf. Yerkes & Dodson, 1908).

Second, we should observe a comparable number of conceptual errors between the immediate evening posttest and the 12-hour posttest the following morning, as the retention interval includes sleep, if there are no substantial circadian effects on the performance measures. If conceptual errors significantly increase from the immediate posttest to the 12-hour posttest (and the increase is statistically comparable to the increase from immediate to 12-hour posttest in Experiment 1), then it suggests that perhaps a simple explanation of "time away from training" or retention interval can explain our results in the first experiment. Motor errors and timing variability, however, which did not show evidence of being reduced by sleep, might increase between the immediate evening posttest and the subsequent morning 12-hour posttest. This pattern of results would support the idea that sleep is serving to selectively consolidate the conceptual – but not specifically motoric – aspects of learning.

METHOD

Participants. Eleven pianists (M = 23.2, SD = 8.1 years old, age range = 18-45, 8 males, 3 females) were recruited to participate in the study. All pianists were able to pass a prescreening measure verifying their ability to sight-read music. Participants were recruited through posting flyers around the University of Chicago campus, as well as sending emails to university piano email groups. Two pianists did not participate in the 12-hour posttest the morning after training, as they either forgot about a prior commitment or overslept. All participants kept a log of their sleep between experimental sessions,

and similar to Experiment 1 all participants reported taking no substances the night before that might alter their sleep.

Materials. We used the same four short pieces of piano music as in Experiment 1. Participants performed the pieces on the same, full-sized, weighted keyboard as in Experiment 1. The keyboard was connected via USB cable to a computer, which recorded MIDI information about the performances including the identity, intensity, and duration (onset and offset) of each note with temporal accuracy of 1 ms. Participants heard the audio output corresponding to each performance through Sennheiser HD 280 headphones at a comfortable listening level (approximately 75 dB SPL). MIDI information was collected using Reason 4.0 (www.propellerheads.se).

Design. All subjects participated in all conditions, with the exception of two participants who did not return for the 12-hour posttest the following morning. The experiment consisted of a training phase, followed by two posttests. Before each session, participants rated their sleepiness on the Stanford Sleepiness Scale (SSS), which is a 7-point subjective measure of sleepiness. Training always occurred in the evening (9pm $\pm$ 1hr). The first posttest immediately followed training, while the second posttest occurred 12 hours after training.

The training paradigm used in the present experiment was identical to the one used in Experiment 1. During the training phase, participants were presented with the sheet music from the four pieces in a pseudo-randomized order (as we always followed a major key piece with a minor key piece or vice versa). Pieces were presented one at a time. Participants could spend up to five minutes learning each piece (see *Design* section of Experiment 1 for details) before being asked to "perform" at a slow tempo of 35 BPM. In order to move onto the next piece, participants had to play at least three successive note-perfect repetitions to ensure that any observed differences in errors were due to experimental manipulations and not to problems in sight-reading or perceptual difficulties. Participants generally made very few note errors during these training "performances," requiring only an average of 3.54 (SD = 0.53) performances to move onto the next piece.

During the posttests, participants performed each of the four pieces three times at a faster tempo. Unlike Experiment 1, the posttest tempos varied between participants to accommodate different levels of performance skill. While it is true that all participants had to possess a certain amount of piano proficiency to be eligible for the study, our participants ranged from amateur pianists who did not regularly practice to professional musicians

who practiced approximately 20 hours per week. Thus, in order to ensure that participants did not reach ceiling performance (committing zero errors), we adjusted the posttest tempos accordingly. These tempos ranged from 50 BPM to 80 BPM depending on performance in the training phase. Regardless of the tempo assigned to a participant, we increased the tempo by 5 BPM between the immediate and 12-hour posttests to offset practice effects that might lead to speeding up, due to increased ease of performing, allowing us to both produce and measure errors. Participants performed with a metronome in order to ensure they were playing at the appropriate tempo.

The order of the pieces was different for each posttest. For example, if a participant received the pieces in the order ABCD during training, they would play the immediate posttest as BCDA, and the 12-hour posttest as CDAB. Before the posttests, participants had the opportunity to warm up (e.g., play scales), but were not allowed to see the pieces, or perform any part of the pieces. Only one participant took advantage of "warming up." At the end of the final session, participant filled out a music experience questionnaire and filled out a sleep log specifically asking them about the quality and duration of their sleep the previous night.

RESULTS

Overall Group Differences. There were no significant differences between the original group of pianists in Experiment 1 and the current control group of pianists in terms of estimated typical playing time from memory versus sheet music [50.5% memory "Experiment 1," 30.9% memory "control," $t(25) = 1.46$, $p = .16$], years of instruction on the piano [11.4 years "Experiment 1," 10.8 years "control," $t(25) = 0.54$, $p = .60$], or self-reported sight-reading ability [3.6 "Experiment 1," 3.7 "control," $t(25) = -0.38$, $p = .71$]. There was, however, a marginally significant difference between the two groups in terms of estimated hours of practice per week [3.5 hours/week "Experiment 1," 7.5 hours/week "control," $t(11.5) = -1.87$, $p = .09$][2], though there were two participants in the current control group (reporting 20 hours/week of practice) that drove this effect.

Errors Immediately After Training. To rule out the possibility that pianists simply perform worse in the evening compared to the mornings, we compared the overall number of substitution errors made by pianists in the current control experiment (trained in the evening) with the overall number of substitution errors made by pianists in Experiment 1(trained in the morning). We included all eleven of our pianists in this analysis. Participants trained in the evening (current control experiment) made an average of 46.27 ($SD = 45.67$) substitution errors in the posttest immediately following training, while participants trained in the morning (Experiment 1) made an average of 29.81 ($SD = 24.11$) substitution errors in the posttest immediately after training. While this pattern of results seems like it might support the notion that pianists simply cannot perform as well in the evenings compared to the mornings, the difference between these two groups was not significant, $t(13.7) = 1.10$, $p = .29$.[3] Thus, we failed to find any conclusive evidence that participants simply made more errors in the evening posttests compared to the morning posttests.

Retention After Sleep. If sleep is actively reducing the errors that manifest over a waking retention interval, then in the current control experiment – where the 12-hour retention interval includes sleep – we would predict no significant increase in the number of conceptual errors made between the immediate, evening posttest and the second, morning posttest. If, however, errors change as a function of elapsed time in the retention interval since active training, then we would predict that conceptual errors would significantly increase from the immediate to the 12-hour posttest, as was observed in Experiment 1. In contrast to conceptual errors, motor errors and CV – which did not show evidence for sleep dependent consolidation – might increase in a comparable manner as was observed in Experiment 1 in a 12-hr retention period. As these analyses required data from both the evening and morning testing sessions, we were only able to include the nine pianists from the current control experiment who completed both testing sessions.

Participants in the current control experiment made an average of 46.67 ($SE = 14.43$) conceptual substitution errors in the evening posttest immediately after training, compared to 45.67 ($SE = 17.60$) conceptual substitution errors in the morning 12-hour posttest following sleep. This difference between the immediate evening and 12-hour morning posttest was not significant, $F(1, 7) = 0.49$, $p = .51$, $\eta_p^2 = .07$. Moreover, the between-subject differences in tempo – which were included as a covariate in the model – did not interact

---

[2] The degrees of freedom were conservatively adjusted as equal variance between the groups was not assumed, Levene's Test: $F(1, 25) = 8.26$, $p < .01$.

[3] The degrees of freedom were conservatively adjusted as equal variance between the groups was not assumed, Levene's Test: $F(1, 25) = 5.49$, $p = .03$.
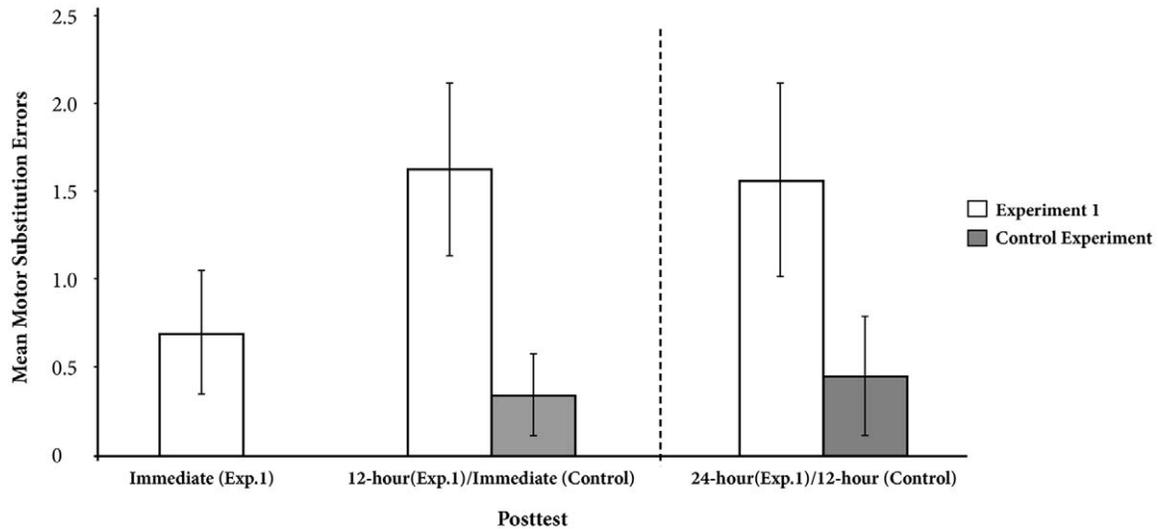
FIGURE 3. Mean motor substitution errors (and SE bars) as a function of posttest for both Experiment 1 and the control experiment. The dotted vertical line represents where sleep occurred.

with the difference in errors between the immediate and 12-hour posttests, suggesting that the fastest tempo participants were not showing a different consolidation pattern from the slowest tempo participants, $F(1, 7) = 0.52, p = .49, \eta_p^2 = .07$. Thus, it appears that participants in the current control experiment made a comparable number of conceptual substitution errors between the immediate and 12-hour posttests. This finding is represented along with the consolidation pattern of Experiment 1 in Figure 2.

To further test this claim (as it rests on a null result), we compared immediate to 12-hour posttest performance of our current control set of participants with the participants from Experiment 1. As we observed a significant increase in conceptual substitution errors from the immediate to the 12-hour waking retention posttest in Experiment 1, we predicted that we should observe a significant interaction between posttest and participant group. To test this hypothesis, we created an ANOVA model with posttests (immediate vs. 12-hour retention) as a within-subjects factor, and participant group (current control experiment vs. Experiment 1) as a between-subjects factor. We found a marginally significant interaction between posttest and participant group, $F(1, 23) = 3.80, p = .06, \eta_p^2 = .14$, providing some evidence that the error patterns for conceptual substitution errors change depending on whether the 12-hour retention interval between the immediate and 12-hour posttests includes sleep.

In terms of motor substitution errors (defined as substitution errors that were removed from the intended target by one semitone and were non-diatonic), participants in the current control experiment made an average of 0.33 ($SE = 0.23$) errors in the evening posttest immediately following training, and 0.44 ($SE = 0.34$) errors in the 12-hour retention posttest the following morning. This increase was in the predicted direction, but was not significant, $t(8) = -0.56, p = .30$. This finding is represented in Figure 3, alongside the motor substitution errors of Experiment 1.[4] However, across all nine participants who completed both posttests, only three motor substitution errors were committed in the immediate posttest, and only four motor substitution errors were committed in the 12-hour posttest, making it difficult to draw any definitive conclusions. Expanding the motor substitution definition to include all semitone errors (as we did in the first experiment), we found that participants made an average of 4.56 ($SE = 1.91$) motor substitution errors in the immediate posttest, and 4.78 ($SE = 2.21$) motor substitution errors in the 12-hour posttest – an increase that was not significant, $t(8) = -0.20, p = .46$.

Given the extremely low number of motor substitution errors made by pianists in the present study, similar

---

[4] The difference in overall motor substitution errors between Experiment 1 and the control experiment was significant in a two-tailed test with equal variances not assumed, $t(20.5) = 2.41, p = .03$. One possible interpretation of this group difference is that the control experiment participants made fewer motor errors because they reported playing the piano marginally longer per week than the Experiment 1 participants, though this interpretation would not explain why the control experiment participants did not also make fewer conceptual errors.
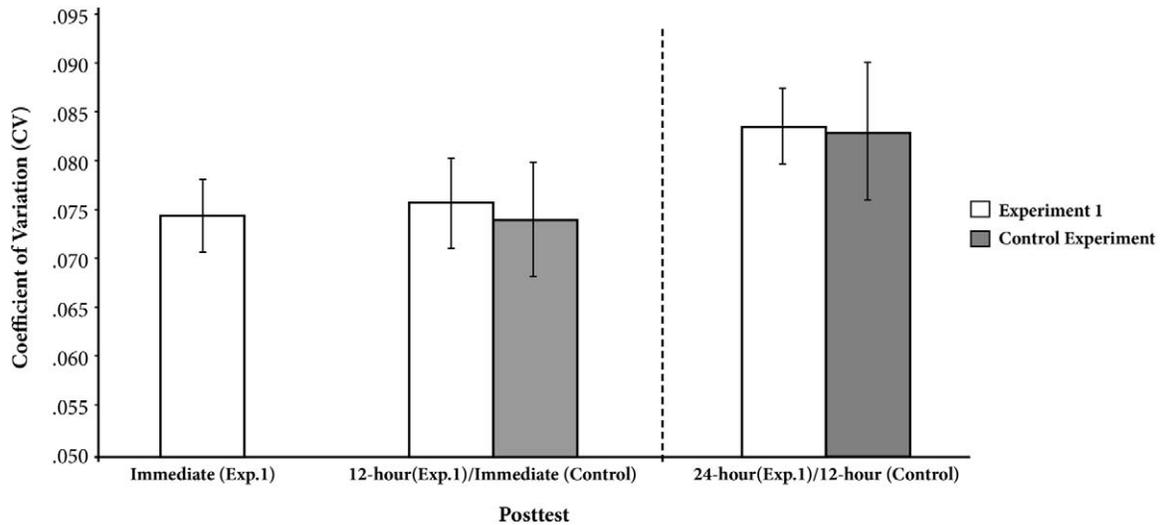
FIGURE 4. Mean coefficient of variation (CV) measures (and SE bars) as a function of posttest for both Experiment 1 and the control experiment. The dotted vertical line represents where sleep occurred.

to Experiment 1 we also examined how timing variability (measured by coefficient of variation, or CV) changed between posttests. In Experiment 1, we observed a significant increase in CV between the 12-hour retention and 24-hour retention posttests – the retention interval including sleep. If sleep is serving to selectively consolidate the conceptual information in music, then we might expect a similar pattern of CV across the immediate and 12-hour retention posttests in the current control experiment; that is, CV might increase between the immediate evening and 12-hour retention morning posttests. The same correction process as Experiment 1 was applied to offset any rounding errors that might occur from the tempo increase of 5 BPM between the immediate and 12-hour posttests. The average CV for the nine participants who completed both posttests was 7.7% (SE = 0.6%) for the immediate posttest after training, and 8.5% (SE = 0.7%) for the 12-hour retention posttest. A one-tailed planned comparison showed that this increase in timing variability was significant, $t(8) = -2.02$, $p = .04$.

As CV is not immune to tempo increases, we also controlled for the between-subject tempo differences by constructing a repeated measures ANOVA with posttest (immediate, 12-hour) as repeated factors and tempo as a covariate. We found some evidence that initial tempo interacted with the sleep consolidation pattern, although the interaction between tempo and posttest was only marginally significant, $F(1, 7) = 3.91$, $p = .09$, $\eta_p^2 = .36$. Nevertheless, 9 of our 11 participants showed an increase in CV from the immediate to the

12-hour posttests larger than would be expected from rounding errors alone, which was significant in a nonparametric test, $X^2(1) = 4.46$, $p = .03$. This pattern of results very closely mirrors the results found in Experiment 1 (see Figure 4), in which CV significantly increased between the 12-hour and 24-hour posttests (the retention interval incorporating sleep).

Playing Styles and Consolidation. Similar to Experiment 1, we correlated the amount of time participants from the current control experiment reported playing from sheet music (versus "playing by ear") with the magnitude of sleep-dependent consolidation for conceptual substitution errors, motor substitution errors, and CV, as well as with the overall number of substitution errors made across all posttests. These results are reported in Table 1. We failed to find any evidence that self-reported playing style was related to the magnitude of sleep-dependent consolidation for conceptual substitution errors, which was different from Experiment 1. This failure to find a relationship between self-reported playing style and the magnitude of sleep dependent consolidation, however, might have been due to low statistical power (indeed, the 95% confidence interval for the correlation coefficient in the current control experiment ($-.72 < \rho < .61$) was large and overlapped with that of Experiment 1 ($.16 < \rho < .85$). Moreover, the combined results from Experiments 1 and 2 still observe a significant correlation between self-reported playing style and the magnitude of sleep-dependent consolidation for conceptual substitution errors ($r = .34$, $n = 25$, $p = .04$).

Motor substitution error consolidation, CV consolidation, and overall substitution errors were not related to playing style, which was consistent with the results from Experiment 1.

Sleep Measures. As in Experiment 1, participants rated their alertness on the Stanford Sleepiness Scale (SSS) prior to each testing session. In service of the specific hypotheses of the current control experiment, we compared participants' evening and morning sleepiness ratings with the evening and morning sleepiness ratings from the participants of Experiment 1. If participants from the current control experiment rated themselves as significantly more or less sleepy than the participants from Experiment 1 during the evening or morning testing sessions, then it could be the case that the observed differences in consolidation between Experiments 1 and the current control experiment are really explainable by general arousal differences between the two groups of participants. This did not seem to be the case – participants in the current control experiment gave themselves an average sleepiness rating of 2.91 ($SD = 1.04$) during the immediate evening posttest, which was not statistically different from the 2.53 ($SD = 0.76$) sleepiness score from the evening 12-hour posttest of Experiment 1, $t(25) = -1.09$, $p = .29$. Moreover, the nine participants of the current control experiment who returned for the 12-hour retention morning posttest gave themselves an average sleepiness score of 3.44 ($SD = 1.33$), which was not statistically different from the immediate morning posttest rating of 3.16 ($SD = 1.18$), $t(23) = -0.56$, $p = .58$, or the 24-hour retention morning posttest rating of 3.31 ($SD = 1.01$), $t(23) = -0.28$, $p = .78$, in Experiment 1. We therefore failed to find any evidence that participants in the current control experiment were different than the participants in Experiment 1 in terms of their self-reported sleepiness across both the evening and morning testing sessions.

An analysis of the nine participants' sleep logs (who completed both testing sessions) showed that participants in the current control experiment reported sleeping an average of 6.26 ($SD = 1.34$) hours during the night between the immediate and 12-hour posttests. This self-reported amount of sleep was not significantly different than the 6.79 ($SD = 1.33$) hours of sleep reported by the participants from Experiment 1, $t(23) = 0.96$, $p = .35$. Self-reported amount of sleep in the current control experiment did not seem to be related to the overall number of substitution errors, the consolidation (operationalized as the reduction of errors from the immediate to the 12-hour posttest) of conceptual substitution errors, the consolidation of motor substitution errors, or

the consolidation of CV. These results are reported in Table 2.

Discussion

Skill acquisition requires learning of various kinds – from the specifics of motor movements, to the integration of relevant sensorimotor information, to the goals and concepts critical for improving performance. Music performance offers a valuable window into how these different kinds of representations develop. In learning a new piece of music, one must learn a complex temporal sequence of motor responses (e.g., hand reaching toward different key presses, violin bows, tightening or loosening of vocal chords), as well as the conceptual structure of music, which dictates particular note constituents of chords or keys. The results from the present study suggest that, in the context of learning piano music, performance knowledge based on a conceptual underpinning of several different melodies may consolidate differently than performance knowledge reflecting motoric information of the musical task.

Specifically, we found evidence that, over the course of learning short piano pieces, conceptual and motor substitution errors significantly increased over a 12-hour waking retention interval, similar to the results of other studies investigating sleep consolidation of skill learning for generalized skills. Following sleep, pianists showed a significant reduction in conceptual errors, while motor errors – as measured by adjacent note substitutions and motor timing variability (CV) – were not significantly reduced after sleep. Moreover, as the control experiment demonstrates, this pattern of results does not seem to be explainable by general time-of-day arousal differences, or "elapsed-time-since-training" (i.e., the absolute amount of time each posttest was removed from first learning the pieces of music). The results from both experiments, taken together, suggest a functional dissociation between the mechanisms of consolidation for motor and conceptual learning.

The dissociation between motor and conceptual information in the current studies is congruent with the findings of recent studies, which have also found functional dissociations in the type of information that is consolidated by sleep. Albouy et al. (2013) demonstrated that in procedural finger-tapping tasks, spatial representations seem to benefit from sleep-dependent consolidation, whereas the specific movement representations do not. In their task, which used an explicit finger-tapping task and a transfer protocol, Albouy et al. (2013) were able to dissociate allocentric (spatial) and egocentric (motor) representations of the learned

sequence, similar to our dissociation of motor and conceptual errors. After a daytime nap, the authors found evidence that sleep strengthens the spatial (but not specifically motor) representation of the sequence, just as we found evidence that sleep strengthens the conceptual (but not specifically motor) representation of music.

In a similar vein, Fenn, Margoliash, and Nusbaum (2013) recently found that sleep provides benefits to generalized – but not rote – learning of synthetic speech. In their paradigm, participants either were trained on a finite set of synthetic speech tokens (rote learning), or a set of synthetic speech tokens that never repeated (generalized learning). The authors found that while performance for both generalized and rote learning degraded over a waking retention interval, generalized learning showed a significant benefit from sleep, whereas rote learning did not.

Although the results of our current study are different from Albouy et al. (2013) and Fenn et al. (2013) in the sense that participants were not presented with a stimulus on which they had not been trained, our dissociation between motor and conceptual representations of music performance similarly suggests that sleep does not provide a blanket enhancement of skill acquisition in a musical context. Indeed, this is similar to the findings of Dumay and Gaskell (2007) that in learning words, sleep did not specifically benefit the learning of the particular trained words, but instead consolidated the integration of the words into the lexicon. In the present study, sleep consolidation of learning the musical pieces means in essence integrating the conceptual representation of the newly learned pieces that differs from prior experience with other musical pieces. Thus, the consolidation of this more conceptual representation abstracted across pieces, resulting in a reduction of conceptual performance errors after sleep.

While motor errors in music performance may seem to relate directly to previous research investigating procedural memory consolidation, it is important to note that neither motor nor conceptual errors can be simply defined as a procedural or declarative task. A conceptual error may be the result of a mistake in declarative learning; however, the expression of this error is nonetheless dependent on procedural learning as well. It is perhaps more informative to consider these errors in terms of *performance* and *competence*.

Motor errors may well be more influenced by the immediate demands of a specific *performance*, while conceptual errors are influenced by musical knowledge or musical *competence* (cf. Chomsky, 1965). Since the instructions for the present task stressed the importance of note accuracy, participants likely focused on learning

the note sequence rather than timing. This potentially explains why we observed the well-documented sleep consolidation pattern in conceptual note errors, but not in temporal variability measures. In support of this idea, Robertson, Pascual-Leone, and Press (2004) have demonstrated that sequence learning only shows offline sleep-dependent benefits if the task is explicitly (not implicitly) framed in the instructions. This suggests that the instructional framing of a given task is highly relevant in predicting whether or not one will observe sleep-dependent performance gains. Applying these results to our current study, we predict that if the instructions of our task were modified to stress highly accurate timing, we might observe a different pattern of results, with timing variability potentially increasing over the course of the day, and then significantly decreasing after sleep. Indeed, this type of result has been observed when the instructions stress accurate timing (e.g., Lewis et al., 2011), providing further evidence that the framing of the task shapes attention, encoding, and ultimately consolidation of procedural knowledge.

Different expectations about the nature of a memory task have long been known to influence memory performance. For example, Morris, Bransford, and Franks (1977) demonstrated that while memory for words belonging to a sentence is better following a semantic task than a rhyme detection task, there was an interaction such that rhyme recognition was better following the rhyme detection task. In the context of learning four new piano pieces, given the emphasis in task framing on note-level performance, musicians appear to focus on the pitches and the patterns of those pitches. This has implications for the consolidation of learning through sleep.

Moreover, the way in which pianists typically practice appears to have implications for the encoding during learning, such that sleep can affect the consolidation of conceptual structure. Specifically, the significant correlation between the reported percentage of time "playing by ear" and the magnitude of conceptual consolidation in Experiment 1 has interesting implications for the ways in which different attentional strategies during encoding may affect consolidation. Future research should assess how different learning strategies (even under identical experimental conditions) can affect the degree to which information is consolidated during sleep. Indeed, it has recently been demonstrated that individual differences in working memory capacity (WMC) predict the degree to which sleep helps consolidate learned information (Fenn & Hambrick, 2012). While we are not implying that "playing by ear" is related to WMC, the measure of playing style reported

here may be related to differences in the way participants attend to and thus approach the learning task. Our findings thus support the idea that that the way in which a memory is initially encoded is crucial in determining how sleep will affect consolidation (Darsaud et al., 2011).

In conclusion, we found evidence that both conceptual and motoric errors in music performance increased over a waking retention interval, but only conceptual errors were significantly reduced after a night of sleep. In contrast, we found no evidence that sleep reduced motor substitution errors or decreased the temporal variability of tone onsets; in fact, the coefficient of variation was significantly greater after sleep compared to performance immediately after training. These findings, moreover, did not seem to be due to potential confounding factors, such as general arousal at various times of day or simple elapsed-time-since-training, as demonstrated by our circadian control experiment. Finally, the magnitude of consolidation for conceptual substitution errors corresponded to pianists' self-reports of playing by ear, suggesting that the manner in which a stimulus is attended to during encoding can significantly affect the extent to which information is consolidated after sleep. These results, taken as a whole, show that sleep does not offer a blanket enhancement in learning complex musical pieces.

Author Note

Correspondence concerning this article should be addressed to Stephen C. Van Hedger, Department of Psychology, University of Chicago, Chicago, IL 60637. E-mail: shedger@uchicago.edu

References

Albouy, G., Fogel, S., Pottiez, H., Nguyen, V. A., Ray, L., Lungu, O., et al. (2013). Daytime sleep enhances consolidation of the spatial but not motoric representation of motor sequence memory. *PLoS ONE 8*(1), e52805. doi:10.1371/journal.pone.0052805

Allen, S. E. (2012). Memory stabilization and enhancement following music practice. *Psychology of Music, 41,* 794-803.

Atienza, M., & Cantero, J. L. (2001). Complex sound processing during human REM sleep by recovering information from long-term memory as revealed by the mismatch negativity (MMN). *Brain Research, 901*, 151-160.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology.* Cambridge, UK: University Press.

Brawn, T. P., Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2008). Consolidation of sensorimotor learning during sleep. *Learning and Memory, 15,* 815-819.

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,* 13977-13982.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Darsaud, A., Dehon, H., Lahl, O., Sterpenich, V., Boly, M., Dang-Vu, T., et al. (2011). Does sleep promote false memories? *Journal of Cognitive Neuroscience, 23,* 26-40.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews, 13*, 309-321.

Duke, R. A., & Davis, C. M. (2006). Procedural memory consolidation in the performance of brief keyboard sequences. *Journal of Research in Music Education, 54,* 111-124.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18,* 35-39.

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia, 49,* 1322-1331.

Fenn, K. M., & Hambrick, D. Z. (2012). Individual differences in working memory capacity predict sleep-dependent memory consolidation. *Journal of Experimental Psychology: General, 141,* 404-410.

Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2013). Sleep restores loss of generalized but not rote learning of synthetic speech. *Cognition, 128,* 280-286.

Fenn, K. M., Nusbaum, H. C., & Margoliash, D. (2003). Consolidation during sleep of perceptual learning of spoken language. *Nature, 425,* 614-616.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17,* 670-674.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Johnson, M. P., Duffy, J. F., Dijk, D. J., Ronda, J. M., Dyal, C. M., & Czeisler, C. A. (1992). Short-term memory, alertness and performance: a reappraisal of their relationship to body temperature. *Journal of Sleep Research, 1,* 24-29.

Jonaitis, E. M., & Saffran, J. R. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33,* 951-968.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., & Sagi, D. (1994). Dependence on REM sleep of overnight perceptual skill. *Science, 265*, 679-682.

KRUMHANSL, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11,* 346-374.

LEWIS, P. A., COUCH, T. J., & WALKER, M. P. (2011). Keeping time in your sleep: Overnight consolidation of temporal rhythm. *Neuropsychologia, 49,* 115-123.

MACKENZIE, C. L., VANEERD, D. L., GRAHAM, E. D., HURON, D. B., & WILLS, B. L. (1986). The effect of tonal structure on rhythm in piano performance. *Music Perception, 4,* 215-225.

MCGAUGH, J. L. (2000). Memory: A century of consolidation. *Science, 287,* 248-251.

MORRIS, C. D., BRANSFORD, J. D., & FRANKS, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.

PALMER, C., & VAN DE SANDE, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 457-470.

PALMER, C., & VAN DE SANDE, C. (1995). Range of planning in music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 947-962.

PALMER, C., MATHIAS, B., & ANDERSON, M. (2012). Sensorimotor mechanisms in music performance: actions that go partially wrong. *Annals of the New York Academy of Sciences, 1252,* 185-191.

PALMER, C., MEYER, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science, 11(1),* 63-68. doi: 10.1111/1467-9280.00216

POHL, P. S., MCDOWD, J. M., FILION, D., RICHARDS, L. G., & STIERS, W. (2006). Implicit learning of a motor skill after mild and moderate stroke. *Clinical Rehabilitation, 20,* 246-253.

REPP, B. H. (1997). Variability of timing in expressive piano performance increases with interval duration. *Psychonomic Bulletin and Review, 4,* 530-534.

ROBERTSON, E. M., PASUAL-LEONE, A., & PRESS, D. Z. (2004). Awareness modifies the skill-learning benefits of sleep. *Current Biology, 14,* 208-212.

SIMMONS, A. L., & DUKE, R. A. (2006). Effects of sleep on performance of a keyboard melody. *Journal of Research in Music Education, 54,* 257-269.

TEMPERLEY, D. (2007). *Music and probability.* Cambridge, MA: MIT Press.

WALKER, M. P, BRAKEFIELD, T., HOBSON, J. A., & STICKGOLD, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425,* 616-620.

WALKER, M. P., & STICKGOLD, R. (2006). Sleep, memory, and plasticity. *Annual Review of Psychology, 57,* 139-166.

YERKES, R. M., & DODSON, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482.