Bucknell University
Bucknell Digital Commons

Honors Theses | Student Theses

2011

Metamemory judgments for familiar and unfamiliar tunes

Amanda Child
Bucknell University

METAMEMORY JUDGMENTS FOR FAMILIAR AND UNFAMILIAR TUNES

by

Amanda E. Child

A Thesis Submitted to the Honors Council

For Honors in Neuroscience

2010-2011

Approved by:

Dr. Andrea Halpern, Adviser

Dr. David Evans, Department Chairperson

Acknowledgements

This project would not have been possible without Prof. Halpern, who tirelessly helped me throughout this process, giving me encouragement and wisdom during every stage. Also, I would like to thank Prof. Rundus and Prof. Jablonski for offering very insightful comments during the thesis defense as well as my family and friends for reading thesis drafts, volunteering their time to pilot test the experiments, and offering moral support. I couldn't have done it without you!

Table of Contents

List of Figures

Abstract

Metamemory is an important skill that allows humans to monitor their own memory abilities; however, little research has concerned what perceptual information influences metamemory judgments. A series of experiments assessed the accuracy of metamemory judgments for music as well as determined if metamemory judgments are affected by ease of processing of musical features. A recognition memory task in conjunction with metamemory judgments (Judgments of Learning, or JOLs) were used to determine actual and predicted memory performance. We found that changing the ease of processing of the volume and timbre of unfamiliar tunes affected metamemory judgments, but not memory performance, for unfamiliar tunes. Manipulating the ease of processing of the timbre and tempo of familiar tunes did not affect metamemory judgments or memory performance although metamemory accuracy on an item-by-item basis was better for familiar tunes as compared to unfamiliar tunes. Thus, metamemory judgments for unfamiliar tunes are more sensitive to ease of processing changes as compared to familiar tunes, suggesting that different types of information are processed in different ways.

Introduction

Explicit Memory

Explicit, or conscious, memories are a fixture of everyday life; they allow people, for example, to remember where they put their keys, recall what their favorite meals are, and recognize their colleagues at work. Since explicit memory is such a critical part of human existence, numerous researchers have studied various aspects of how explicit memory functions, such as how it is affected by aging or brain injury (Fleishman, Wilson, Gabrieli, Bienias, & Bennett, 2004; Yeates & Enrile, 2005), which brain areas are involved in explicit memory formation (Voss & Paller, 2008), and what memory strategies maximize the probability that a person will be able to remember something at a later time (Justice & Weaver-McDougall, 1989).

One particularly interesting area of explicit memory research explores why some things we encounter in real life are better remembered than others; for example, most people are pretty good at recognizing people they have met before, but it is not uncommon for those same people to have difficulty remembering their acquaintances' names. In order to approach a question such as this, experimenters may test participants' memories using a recall task. In this study design, participants are first presented with a list of items; for example, participants may be shown a series of photographs and told the name of the person in the photograph. After a certain amount of time, they are asked to retrieve the name of the person in the photograph (McGillivray & Castel, 2010). However, when stimuli are more difficult to learn or it is impractical to expect participants to reproduce the stimuli (e.g. when studying memory of paintings), experimenters typically use recognition memory tasks. During this type of experiment, after being presented with the target information, participants are presented with a series of items and asked to distinguish which items they had already been exposed to from which information was novel. For example, in an experiment by Vogt and Magnussen (2007), participants saw a list of pictures. Later, they viewed a list of pairs of pictures containing one picture they had seen earlier and one picture they had not seen earlier and were asked to identify which picture they recognized.

Metacognition and Metamemory

Humans also have the ability to monitor their cognitive abilities, a skill that is called *metacognition*. As an example, when approaching a problem, people first determine which of their problem-solving strategies would likely be the most effective and then implement the strategy (Dunlosky & Metcalfe, 2009). One important type of metacognition is called *metamemory*, or the monitoring of memory abilities. For example, as students study for exams, they periodically judge their memory for the course material and typically stop studying when they feel that their knowledge of the material is reliable. In general, metamemory judgments are accurate in that they are fairly good predictors of memory performance. For example, Bunnel, Baken, & Richards-Ward (1999) found that metamemory accuracy in young adults was high when they were presented with a list of words, asked to estimate how many words they would be able to recall, and then recalled the items. Interestingly, the accuracy of these judgments, past young adulthood, remains constant as memory declines with age (Hertzog, Dixon, & Hultsch, 1990).

However, metamemory judgments are not always accurate predictors of actual memory performance. For example, in an experiment exploring the effect of text difficulty on metamemory (Weaver & Bryant, 1995), participants were presented with four passages and asked to rate their level of comprehension of each passage. They subsequently responded to a series of multiple-choice questions that tested their memory for the passage. The authors found that participants' metamemory judgments were well calibrated with texts of intermediate difficulty and less well calibrated with easy and difficult texts (Weaver & Bryant, 1995). Another experiment (Maki, Willmon, & Pietan, 2009) explored the accuracy of college students' metamemory judgments for different types of tests by presenting them with a series of short passages, asking them to estimate how well they would do on a multiple-choice, essay, or recall test for those passages and subsequently giving them one of those three types of test for each passage. Participants were found to have more accurate metamemory judgments for multiple-choice tests as compared to essay and recall tests (Maki, Willmon, & Pietan, 2009), suggesting that metamemory accuracy is influenced by the specific task being performed.

Illusory Learning

The dissociation between metamemory judgments and memory performance is also seen in *illusory learning*, which describes a situation in which people believe that memory performance will be influenced by the value of a particular feature, when in fact performance is actually unchanged. For example, Rhodes and Castel (2008, 2009) found that whereas both text size and volume affected how well participants judged they would remember visually or aurally presented words during the test phase (Judgment of

Learning, or JOL), their memory performance was not affected by either of those factors. In other words, participants thought they were more likely to remember text that was presented in a larger font or words that were spoken in a louder volume. However, during the recall task, they were no more likely to remember large or loud words as compared to soft or small words. This is a particularly interesting finding because it suggests that metamemory judgments may be influenced by the *ease of processing* of the stimuli (Rhodes & Castel, 2008; Rhodes & Castel, 2009); in other words, information that seemed easy to process (loud or large words) was thought to be better learned as compared to information that seemed hard to process.

Music Memory

Music, like memory, is an integral part of everyday life. As a society, Americans spend millions of dollars each year on concert tickets and music purchases, and in addition to the number of people that have previous music experience, 37% of survey respondents currently played a musical instrument in 2003 (Lyons, 2003). Furthermore, many people memorize music at some point in their lives, whether they would like to sing the correct melody and lyrics in the shower or in their solo on Broadway. However, memory for music is an understudied area, relative to the study of how we remember written or spoken words; to illustrate, a literature search on PsycInfo for “music memory” generated 42 articles while a search for “language memory” generated 534 hits. While there are similarities between language and music, including the use of a metrical structure and pitch contour, music also differs from language; for example, it requires the use of distinct pitches (Jackendoff, 2009). Furthermore, words have distinct meanings while music does not; for example, the word "plum" universally refers to a round, purple fruit, but the tune *Für Elise* is not necessarily associated with anything outside of itself. This suggests that memory for music is likely distinct from language memory and underscores the need to study music memory specifically.

Some experimenters have already made important progress in understanding memory for both familiar and novel melodies. For example, in regards to familiar tunes, researchers have discovered that people are particularly good at accumulating a large library of songs they can recognize or reproduce (Halpern & Bartlett, 2010), and these music memories can last for decades (Schulkind, Hennis, & Rubin, 1999), even withstanding the damages of early-stage Alzheimer's disease (Bartlett, Halpern, & Dowling, 1995). In contrast to this apparent affinity of memory for familiar songs, research has found that memory for novel instrumental melodies is notoriously poor, particularly when compared to memory for other types of novel information; it is not uncommon for performance on a novel music recognition task using 40 tunes to be only modestly above chance (Halpern & Müllensiefen, 2008). In comparison, in similar tasks using 400 novel paintings or up to 48 words, participants' performance was significantly above chance (Vogt & Magnussen, 2007; Tulving & Pearlstone, 1966). In general, it appears as though one repetition is not sufficient to reliably encode music into long-term memory; however, once music is successfully encoded, it seems that the memory will be fairly permanent.

Researchers have also studied similarities and differences between music memory and language memory. One example of their differences is illustrated by an experiment by Groussard et al. (2010), who found that music and verbal memory each activate distinct areas of the left temporal cortex. In addition, an experiment by Warker and Halpern (2005) failed to find that more deeply encoding music leads to better memory performance (Level of Processing, or LOP), an effect that is often found in other domains including verbal memory (Bentin, Moscovitch, & Mirhod, 1998). However, music memory and verbal memory share some characteristics. For example, explicit memory for music and verbal information declines with age (Chandler et al., 2004; Gaudreau & Peretz, 1999). Also, neuroimaging studies have revealed an overlap in the brain areas activated during music and verbal memory retrieval (Groussard et al., 2010).

Music Metamemory

Despite these advancements in the study of music memory, very little progress has been made in the study of music metamemory. In a real-life context, music metamemory allows people to assess the strength of their memory for music prior to testing the memory, which could be a concerto solo or a karaoke solo. Studies exploring music metamemory accuracy could provide helpful insight that could allow performers to better assess their memory. Furthermore, these studies may provide understanding into how metamemory judgments are actually made; because metamemory judgments are simply an estimation of future memory performance, people likely use cues from the information to guess the likelihood it will be remembered later. However, it is unclear exactly what those cues are. Since music has been shown to be distinct in some respects from other forms of information, such as words, differences in metamemory performance accuracy in music and other items may provide insight into what cues determine these judgments.

The few music metamemory experiments that have been conducted compared feeling of knowing (FOK) judgments (a type of metamemory judgment) and memory performance for tunes (Korenman & Peynircioglu, 2004; Peynirciolgu, Rabinovitz, & Thompson, 2008; Peynirciolgu, Tekcan, Wagner, Baxter, & Shaffer, 1998). In one experiment (Peynircioglu, Rabinovitz, & Thompson, 2008), participants were presented with a series of familiar melodies, song titles, or song lyrics and were asked to recall another piece of information (melody, title, lyrics) about the tune. If recall failed, they were then asked to judge the likelihood they would recognize the correct response if given a short list of possible answers (FOK judgment). Then, they were provided with a few possible options and chose the one they recognized. This experiment found that melody and title cues elicited higher FOK ratings than lyric cues, which the authors hypothesized was because melodies and titles felt more familiar; however, accuracy was the same for all conditions (Peynircioglu, Rabinovitz, & Thompson, 2008). Another experiment (Korenman & Peynircioglu, 2004) used a similar procedure where melodies of varying familiarity were paired with animal-name titles (i.e. “dog”). After participants were cued by the melody or the title, they were asked to recall the other piece of information or, if recall failed, to give a FOK judgment followed by a recognition task. The authors found that FOK judgments were more accurate for familiar tunes, although they were not higher for familiar tunes (Korenman & Peynircioglu, 2004).

This area of research is finding that metamemory for music, like in other domains, is fairly accurate, although accuracy varies with the familiarity of the tunes. However, the distinctions between memory for music and verbal information, which is structurally the closest type of information to music, strongly suggest that music metamemory may not be identical to verbal or other types of metamemory. Given the significance of music metamemory to many people as well as the potential to learn more about how the brain makes metamemory judgments, it is important to conduct more experiments in order to better understand how music metamemory judgments are made.

Current Experiments

In an effort to improve knowledge regarding the relationship between music memory and metamemory, we conducted three experiments that compared JOLs and recognition memory performance for a list of tunes. In Experiments 1 and 2, we compared JOLs and memory performance for unfamiliar tunes that varied in ease of processing of timbre and volume. In Experiment 3, the JOLs and memory performance were compared for familiar tunes that varied in ease of processing of timbre and tempo. Timbre, volume, and tempo were chosen because varying these features does not change the identity of the tune, but would likely change the ease of processing of the tune. For example, “Happy Birthday” is the same tune whether it is played in a organ timbre, in a loud volume, or in a fast tempo, but we hypothesize that a familiar organ timbre, loud volume, and standard tempo are easier to process as compared to an unfamiliar recorder timbre, soft volume, and fast or slow tempi.

For all three experiments, these musical features were varied in such a way that we predicted would influence the ease of processing of the tune. We hypothesized that loud tunes would be easier to process as compared to soft tunes simply because they are easier to hear. Also, we predicted that tunes in a more familiar timbre would be easier to process as compared to tunes in a less familiar timbre. This is supported by previous research that suggests that metamemory judgments can be affected by the *familiarity* of a feature value (Maki, R.H. & Serra, M., 1992), raising the possibility that familiarity may be related to ease of processing. With this in mind, we used an organ timbre (more familiar, easier to process) and a recorder timbre (less familiar, less easy to process) for Experiments 1 and 2. We also predicted that familiar tunes played in their standard tempo would be easier to process than tunes played in a non-standard tempo (fast or slow) because people generally hear tunes played in their standard tempo, making this tune version more familiar and easier to process as compared to fast or slow versions.

We hypothesized that varying the ease of processing of certain tune features would induce the illusory learning effect in participants. In other words, tunes with features that are easier to process would lead participants to believe they are more likely to recognize the tune during a memory test, but actual memory performance would not be affected by how easy the tune is to process. In order to test this hypothesis, we varied the ease of processing of the volume, timbre, and tempo of the tunes throughout the three experiments. In each experiment, participants heard a series of tunes with varying characteristics and made a JOL judgment following each tune. JOL judgments allowed us to assess predictions of future memory performance and do not require recall of target information, unlike feeling-of-knowing judgments. Because music memory for novel melodies is so poor and participants have varying abilities to sing on pitch, a recall task would be nearly impossible for participants to complete. Later, they heard another series of tunes, half of which they had heard earlier and half of which were new. Following each tune, they indicated whether or not they recognized the tune. At the conclusion of the experiments, JOL judgments and recognition memory performance were compared for tunes with low and high levels of processing as well as different characteristics (volume, timbre, tempo).

Experiment 1

Methods

Participants

The 29 participants were undergraduate students at least 18 years of age recruited on a volunteer basis from the PSYC100 subject pool. Participants' musical experience ranged from 0 to 10 years with a mean of 3.2 years (*SE* = 0.7).

Materials

Eighty unfamiliar tunes were used in this experiment. Unfamiliar tunes were defined as either little-known classical melodies or equally pleasant tunes constructed for previous use in this laboratory. A subset of 24 tunes were created by conserving the note values, intervals, rhythms, and tempi of familiar tunes (such as *Twinkle, Twinkle*) but were not recognizable. Each tune was between 3 and 11 s long and consisted of a single melody line composed in a major key. All tunes were generated in two volumes (soft, 75 dB, and loud, 95 dB) and timbres ("organ" and "recorder"). Previous pilot testing ensured that these timbres were equally pleasant to listen to, but varied in familiarity (organ = more familiar, recorder = less familiar). A MIDI synthesizer (Yamaha PSR-500) and Cakewalk Pro Audio software were used to synthesize the “organ” and “recorder” versions of the tunes. Audacity computer software was used to adjust the volumes of the soft and loud versions as well as to convert the MIDI files into .wav files. All soft and loud tunes were synthesized in a neutral “piano” timbre on the MIDI synthesizer and all organ and recorder tunes were synthesized in a neutral volume (85 dB).

Procedure

Experiment 1 was divided into a study phase and a test phase (Figure 1). During the study phase, each participant heard 20 songs in a Volume block and 20 songs in a Timbre block, resulting in 40 songs presented during the entirety of the study phase. In each block, half of the tunes were loud (or organ) and half were soft (or recorder), depending on the block. The order of the tunes within each block was randomized and no more than four tunes with the same characteristic (i.e. soft, recorder, …) were played in a row. After each tune was played, participants were given 4 s to judge on a scale from 0%-100% the likelihood they would recognize the tune during a later recognition memory task (Judgment of Learning, or JOL).

Following the study phase, participants took about 5 min to fill out a musical background questionnaire that detailed their musical experience. Participants then began the test phase. Each participant heard 40 songs in a Volume block and 40 songs in a Timbre block. Within each block, half of the tunes had been played during the study phase (“old”) and half were new to the participants (“new”). The old tunes reappeared in the test phase with the same characteristics they had had in the study phase (i.e. a song heard in a recorder timbre in the study phase would be played in a recorder timbre in the test phase as well). Also, half of the new tunes in each block were loud (or organ) tunes and half were soft (or recorder) tunes. The order of the tunes within each block was randomized and no more than four tunes with the same characteristic (i.e. soft, recorder, …) were played in a row.

Following each tune, participants were given 5 s to indicate whether or not they believed they had heard the tune during the study phase as well as how confident they were in their response on a scale of 1-6, 1 meaning they were certain the song was new and 6 meaning they were certain the song was old.

To avoid order effects, half of the participants were presented with the Volume block first and the other half was presented with the Timbre block first in both the study and test phases. In addition, each version of each tune (soft, loud, organ, recorder) was presented approximately equally often across participants. Furthermore, each tune version appeared approximately equally often as an "old" and "new" tune, which also controlled for effects of tune length. Overall, eight versions of the experiment were used to achieve this counterbalancing.

Results

Before performing statistical analyses, we excluded the data of four subjects whose AUC scores were below chance, suggesting that they misinterpreted the recognition memory scale. To assess metamemory judgments, average JOL ratings were calculated and an ANOVA was performed. Overall average JOL ratings ranged from 0.50

to 0.58 (Figure 2), which is approximately midway in the JOL scale that ranged from 0 to 1. JOL means and standard errors (Figure 2) suggest that there was no statistically significant difference between the JOLs assigned to loud vs. soft or organ vs. recorder songs. This is confirmed by an ANOVA comparing the following factors: feature (volume or timbre) and ease of processing (high or low). The ANOVA analysis revealed no main effect of feature (volume or timbre) ($F(1,23) = 3.33$, $p = 0.081$), no main effect of level of processing (high or low) ($F(1,23) = 2.00$, $p = 0.171$), and no interaction between the two variables ($F(1,23) = 2.00$, $p = 0.171$).

To assess memory performance, confidence ratings were converted to area under the curve (AUC) scores, which were derived from the signal detection theory, for each participant. This score revealed how good a participant was at discriminating between old and new items. A score of 0.50 indicates the participant was performing at chance, suggesting she or he did not recognize the old items, whereas a score of 1.0 indicates the participant's discrimination between old and new items was perfect and highly confident, suggesting her or his memory for old items was perfect (Halpern & Müllensiefen, 2008). Overall average AUC scores ranged from 0.68 to 0.71, which is significantly above chance performance. The means of the AUC scores were not statistically different between loud, soft, organ, and recorder songs (Figure 2). This is confirmed by an ANOVA that revealed no main effect of feature (volume or timbre) ($F(1,23) = 0.00$, $p = 0.979$), no main effect of ease of processing (high or low) ($F(1,23) = 0.01$, $p = 0.328$), and no interaction between the two variables ($F(1,23) = 0.03$, $p = 0.233$). A 3-way

ANOVA also revealed no interaction among measure (JOL or AUC), feature (volume or timbre), and ease of processing (high or low) ($F(1,23) = 0.04$, $p = 0.848$).

Kruskal-Goodman gamma correlations were also used for each participant to determine if metamemory judgments were predictive of memory performance on an item-by-item basis (Rhodes & Castel, 2008; Rhodes & Castel, 2009). Gamma correlations are commonly used in metamemory literature and are ideal for assessing the relationship between non-normalized variables. The predictor was the JOL judgment and the outcome was the confidence score. Only old items were used for this correlation because items had to have both JOL judgments and recognition memory scores; as a result, a positive gamma correlation indicates that increasing metamemory judgments are predictive of more confident recognition memory scores. The average correlations for their responses overall (*gamma* = 0.15, 95% CI [0.09, 0.20]) and for their responses on volume items (*gamma* = 0.17, 95% CI [0.09, 0.26]) differed reliably from zero ($t(23) = 4.65$; $t(23) = 3.49$); however, the average correlations for timbre items (*gamma* = 0.08, 95% CI [-0.01, 0.17]) did not differ from zero ($t(23) = 1.45$). This means that metamemory judgments were significantly correlated to memory performance for all items as well as for volume items, but not for timbre items.

Discussion

Overall, the results from Experiment 1 suggest that neither metamemory judgments nor memory performance for music are influenced by ease of processing. This result did not support our hypothesis that metamemory judgments would be higher for easier-to-process tunes as compared to harder-to-process tunes despite no difference in actual memory performance. The findings were not consistent with previous experiments exploring the illusory learning phenomenon (Rhodes & Castel, 2008; Rhodes & Castel, 2009); thus, it is possible that ease of processing does not influence music metamemory judgments. However, there was a small, but significant, correlation between participants' metamemory judgments and their memory performance on an item-by-item basis overall as well as for volume items, suggesting that JOLs for music are a somewhat accurate predictor of future recognition memory performance, particularly when ease of processing of volume varies. This finding was consistent with previous music metamemory studies (Korenman & Peynircioglu, 2004; Peynirciolgu et al., 1998; Peynircioglu, Rabinovitz, & Thompson, 2008), which found positive gamma correlations between FOK judgments and memory performance. It is not clear why the gamma correlation for timbre items was not significant. However, all of the correlations were fairly weak, suggesting that while participants were slightly better calibrated for volume items as compared to timbre items, the effect was not strong.

Furthermore, we had a few concerns regarding the procedure. First, only three of the participants used the entire range of the JOL scale and six participants used less than half of the scale (ex. 20%-60%). It is possible that JOL judgments may have been affected by ease of processing, but the effects were not noticeable in this experiment because a majority of participants used a limited range of the JOL scale. Also, although memory performance was respectable and significantly above chance, it would ideally be higher.

Experiment 2

Participants showed in Experiment 1 that they could successfully recognize the novel tunes, despite the generally difficulty of learning unfamiliar music. We decided to follow up on these results with a subsequent experiment using a modified version of the Experiment 1 procedure. For Experiment 2, we decided that encouraging the participants to use the entire range of the scale as well as decreasing the difficulty of the memory task would allow us to more accurately assess whether or not our hypothesis was supported. Also the procedure order was changed so the study and test phases were closer together in time, hopefully improving memory performance. Also, in Experiment 1, the JOL scale ranged from 0% to 100%, which participants may interpret as ranging from "I will definitely not remember this tune" to "I will definitely remember this tune." In order to encourage participants to use the entire range of the scale, the JOL scale ranged from 1 to 10 in experiment 2. Instructions specified that a response of 1 meant that participants thought it was "highly unlikely" they would recognize the tune and 10 meant that they thought it was "highly likely" they would recognize the tune. Furthermore, explicit instructions to use the entirety of the JOL scale throughout the study sessions were added. Finally, brief buffer tasks were added in between the study and test phases to eliminate the possibility of a recency bias in their memory performance.

Methods

Participants

The 21 participants were undergraduate students at least 18 years of age recruited on a volunteer basis from the PSYC100 subject pool. Participants' musical experience ranged from 0 to 12 years of private lessons with a mean of 3.6 years (*SE* = 0.8).

Materials

The tunes used in experiment 1 were also used in experiment 2 in the same Volume and Timbre conditions. Counterbalancing was achieved as described in Experiment 1.

Procedure

In Experiment 2, each study and test phase contained the same songs presented in the same order and with the same characteristics as in experiment 1. However, instead of being presented with both study phases followed by a brief break and both test phases, participants were presented with a Volume or Timbre block, a musical background questionnaire, and the remaining volume or timbre block (Figure 3). Each block consisted of a study phase, a 2-minute distractor task (listing the United States or state capitals), and a test phase.

Results

We excluded the data of two subjects whose AUC scores were below chance, suggesting they misinterpreted the recognition memory scale. To assess metamemory judgments, we calculated average JOL ratings and performed an ANOVA. Overall average JOL ratings ranged from 0.50 to 0.60 (Figure 4), which is approximately midway in the JOL scale that ranged from 0 to 1. JOL means and standard errors (Figure 4) suggested differences between the JOL ratings given to loud and soft songs as well as organ and recorder songs. An ANOVA analysis compared the following factors: feature (volume or timbre) and level of processing (high or low). The ANOVA revealed no main effect of feature (volume or timbre) ($F(1,20) = 0.10$, $p = 0.71$), a main effect of ease of processing (high EOP: $M = 0.59$, $SE = 0.02$; low EOP: $M = 0.50$, $SE = 0.02$; $F(1,20) = 17.58$, $p = 0.00$), and no interaction between the two variables ($F(1,20) = 0.60$, $p = 0.45$).

To assess memory performance, confidence ratings were converted to AUC scores for each participant. Overall, average AUC scores ranged from 0.67 to 0.74, which is significantly above chance (chance AUC score = 0.50). The means of the AUC scores were not statistically different between organ and recorder tunes, but there did appear to be a difference between the means of the loud and soft AUC scores (Figure 4). However, an ANOVA found no main effect of feature (volume or timbre) ($F(1,20) = 1.00$, $p = 0.33$), no main effect of ease of processing (high or low) ($F(1,20) = 3.00$, $p = 1.00$), and no interaction between the two variables ($F(1,20) = 2.00$, $p = 0.17$). Furthermore, a 3-way ANOVA revealed no interaction between measure (JOL or AUC), feature (volume or timbre), and ease of processing (high or low) ($F(1,20) = 0.15$, $p = 0.71$).

In order to assess the role of musical experience in both metamemory judgments and memory performance, we performed ANCOVAs using musical experience (as indexed by years of music lessons) as a covariate. After removing musical experience as a covariate, several new effects emerged. For JOLs, we found a significant interaction between attribute and ease of processing for JOLs ($F(1,20) = 4.42$, $p = 0.05$) and for

AUC scores, we found main effects of both attribute ($F(1,20) = 7.71$, $p = 0.02$) and ease of processing ($F(1,20) = 6.65$, $p = 0.02$). Also, the ANCOVA eliminated a main effect of ease of processing on JOL judgments that we had seen with the ANOVA.

Because musical experience appeared to affect JOL judgments and AUC scores, we divided the participants by the median split of musical experience into two groups: low musical experience with an average of 0.70 years of experience ($SE = 0.30$) and high musical experience with an average of 6.27 years of experience ($SE = 0.76$) (Figures 5 & 6). We then performed additional ANOVAs using musical experience group as a between-subjects factor. There was no interaction among attribute, ease of processing, and musical experience group ($F(1,20) = 1.97$, $p = 0.18$) for JOLs, suggesting that musical experience does not significantly affect JOL judgments, but there was an interaction among attribute, ease of processing, and musical experience group ($F(1,20) = 10.58$, $p = 0.004$) for AUC scores.

In order to assess the effect of musical experience group on AUC scores, we performed separate ANOVAs for the AUC scores of participants with low and high musical experience. For the low experience group, the mean AUC scores for easy-to-process tunes appeared to be significantly higher than the mean AUC scores for harder-to-process tunes (Figure 6), which was confirmed by a main effect of ease of processing ($F(1,9) = 5.41$, $p = 0.05$). We found no main effect of attribute ($F(1,9) = 2.61$, $p = 0.14$) but a significant interaction between ease of processing and attribute ($F(1,9) = 15.70$, $p = 0.003$). This appears to be localized to volume; that is, loud tunes were better remembered compared to soft tunes whereas memory performance was equal for organ and recorder tunes. For participants with high levels of musical experience (Figure 6), we found no effects that were close to being significant.

The average Kruskal-Goodman gamma correlations were weak, but differed significantly from zero, for their responses overall (*gamma* = 0.13, 95% CI [0.07, 0.19]; $t(20) = 3.70$), for their responses on volume items (*gamma* = 0.14, 95% CI [0.07, 0.22]; $t(20) = 3.21$)), and for their responses on timbre items (*gamma* = 0.11, 95% CI [0.03, 0.18]; $t(20) = 2.33$).

Discussion

The results from Experiment 2 suggest that the procedural adjustments intended to improve memory performance as well as use of the JOL scale were successful. Overall memory performance improved from Experiment 1 ($M = 0.68$, $SE = 0.01$) to Experiment 2 ($M = 0.72$, $SE = 0.01$) ($t(88) = 1.72$), suggesting that putting the study and test phases together for each condition was mildly successful at boosting memory. Also, a majority of the participants used the entirety of the JOL scale and no participants used less than 80% of the scale (i.e. 2 to 9).

The results also showed a main effect of ease of processing on JOL judgments, demonstrating that participants estimated they were more likely to remember tunes with easier to process features (loud, organ) as compared to tunes with harder to process features (soft, recorder). At the same time, there was no main effect of ease of processing for AUC scores. Overall, these results support our hypothesis that metamemory judgments, but not memory performance, for music items are influenced by ease of processing, which is consistent with previous studies exploring the illusory learning phenomenon in other domains (Rhodes & Castel, 2008; Rhodes & Castel, 2009). Furthermore, gamma correlations were small, but significant overall as well as for volume and timbre items specifically, suggesting that metamemory performance is modestly well calibrated to actual memory performance for novel tunes.

After dividing participants into groups based on their musical experience, we found that participants with low levels of musical experience had significantly better memory performance for easier to process vs. harder to process tunes whereas the memory performance of participants with high levels of musical experience were not affected by ease of processing. In conjunction with the overall main effect of ease of processing, we conclude that participants with low levels of musical experience correctly predicted they would have better memory performance for items with easier to process features. In contrast, participants with high levels of musical experience also predicted they would have better memory performance for items with easier to process features; however, their memory performance was equal for easier and harder to process items. This suggests that the illusory learning phenomenon primarily occurred in more well-trained people. Apparently, encoding of new music is not influenced by ease of processing for people with more musical experience, although they, curiously, are unaware of this fact.

Experiment 3

Experiment 3 was conducted to determine if the illusory learning phenomenon occurs with tunes that vary in timbre or tempo. In this experiment, we decided to use tunes that participants were already familiar with, which are more analogous to the common words used in previous experiments exploring the illusory learning phenomenon (Rhodes & Castel, 2008; Rhodes & Castel, 2009). Furthermore, we decided to vary timbre and tempo in Experiment 3. We chose tempo as a new feature because it is one of the few musical features that can be varied in such a way that may affect ease of processing, but does not distort the tune. In connection with the ease of processing theory, we hypothesized that tunes played in their standard tempo would be easier to process because people are more familiar with the tune in that tempo. Conversely, tunes played at a faster or slower rate than usual would likely be more difficult to process.

Methods

Participants

The 28 participants were undergraduate students at least 18 years of age recruited on a volunteer basis from the PSYC100 subject pool. Participants' musical experience ranged from 0 to 12 years with a mean of 3.4 years (*SE* = 0.7).

Materials

Sixty familiar verbal tunes (i.e. tunes with well-known lyrics, such as *America the Beautiful* and *Twinkle, Twinkle*) as well as 20 familiar nonverbal tunes (such as the theme songs from *Indiana Jones* and *Looney Tunes*) were used in this experiment. Previous pilot testing of the tunes at their baseline tempo ensured that tunes were familiar to a majority of college-age students. Tunes ranged from 3 to 14 s (including slow and fast versions). Each tune was generated in a "recorder" and "organ" timbre (as described in experiment 1) as well as in fast, regular, and slow speeds in a "piano" timbre. Fast tunes and slow tunes were 20% faster and 20% slower, respectively, as compared to baseline tempo, which created a noticeable, but not distorting, change in tempo. These changes were created using Cakewalk, a program that could change the tempo without changing the pitch of the tune. Counterbalancing was achieved as described in Experiment 1.

Procedure

The overall procedure of experiment 3 was the same as that of experiment 2 (Figure 7). During both the study and test phases of the tempo block, half of the songs were presented in a familiar tempo (regular) and half were presented in an unfamiliar tempo (fast or slow) with equal numbers of fast and slow songs occurring in both phases.

Results

We excluded the data of one subject whose AUC scores were below chance, suggesting he or she misinterpreted the recognition memory scale. To assess metamemory judgments, we calculated average JOL ratings and performed an ANOVA. Overall average JOL ratings ranged from 0.65 to 0.69 (Figure 8), which is above midway in the JOL scale that ranged from 0 to 1. JOL means and standard errors (Figure 8) suggest that JOL ratings given to tunes played in a standard tempo, a non-standard tempo, an organ timbre, or a recorder timbre did not significantly differ from each other. This was confirmed by an ANOVA analysis that revealed no main effect of feature (volume or timbre) ($F(1,26) = 0.16$, $p = 0.95$), no main effect of ease of processing (high or low) ($F(1,26) = 0.07$, $p = 0.79$), and no interaction between the two variables ($F(1,26) = 1.17$, $p = 0.20$). Also, an additional ANOVA analysis confirmed that there was no significant effect of non-standard tempo (fast or slow) on JOL judgments ($F = 0.92$, $p = 0.35$).

To assess memory performance, confidence ratings were converted to AUC scores for each participant. Average AUC scores ranged from 0.85 to 0.89, which is well above chance (Figure 8). The means of the AUC scores were not statistically different between tunes presented in standard and non-standard tempi or between organ and recorder tunes; however, there does appear to be a statistical difference between the means of the fast ($M = 0.82$, $SE = 0.03$) and slow AUC scores ($M = 0.91$, $SE = 0.02$). An ANOVA found no main effect of feature (tempo or timbre) ($F(1,27) = 0.16$, $p = 0.69$), no main effect of ease of processing (high or low) ($F(1,27) = 0.07$, $p = 0.79$), and no interaction between the two variables ($F(1,27) = 1.17$, $p = 0.29$). However, due to an apparent difference between the mean AUC scores for fast and slow tunes, we conducted an additional ANOVA that revealed a significant effect of non-standard tempo (slow greater than fast) on AUC score ($F(1,27) = 6.25$, $p = 0.02$). Furthermore, in order to assess the effect of musical experience on metamemory judgments and memory performance, we conducted ANCOVAs using years of musical experience as a covariate and found that this did not change any of the previous results.

The average Kruskal-Goodman gamma correlations were strong and significantly differed from zero for participants' overall responses (*gamma* = 0.45, 95% CI [0.39, 0.51]; $t(27) = 12.27$), for their responses on tempo items (*gamma* = 0.52, 95% CI [0.43, 0.60]; $t(27) = 10.54$), or for their responses on timbre items (*gamma* = 0.39, 95% CI [0.30, 0.48]; $t(27) = 7.60$).

Discussion

In Experiment 3, both AUC scores as well as Kruskal-Goodman gamma correlations were very high, suggesting that both memory performance and item-by-item calibration of metamemory judgments to memory performance were very good for familiar tunes. The fact that memory performance improved with the use of familiar tunes as compared to unfamiliar tunes was consistent with previous literature comparing recognition memory for the two types of tunes (Bartlett, Halpern, & Dowling, 1995; Dowling, Bartlett, & Halpern, 2008). Also, the high correlations between metamemory judgments and memory performance were similar to those seen with familiar tunes in the music metamemory studies (Peynircioglu et al., 1998; Korenman & Peynircioglu, 2004).

Furthermore, the lack of main effects of feature and ease of processing in the ANOVAs analyzing JOL judgments and AUC scores both in general as well as when participants were divided into groups based on musical experience demonstrate that ease of processing failed to affect either metamemory judgments or memory performance, suggesting that participants' metamemory judgments were well calibrated to memory performance on a general level as well regardless of musical experience. Although previous studies that found that ease of processing did affect metacognitive judgments for texts (Rawson & Dunlosky, 2002; Dunlosky, Baker, Rawson, & Hertzog, 2006) or for written and spoken words (Rhodes & Castel, 2008; Rhodes & Castel, 2009), these studies were testing unfamiliar information; we are not aware of studies that tested whether ease of processing affected metamemory or memory of familiar information.

In addition, the significant effect of non-standard tempo on AUC score in conjunction with the lack of effect of non-standard tempo on JOL judgments suggests that memory performance was better for slow tunes as compared to fast tunes as well as that the participants were unaware of this memory effect.

General Discussion

Overall, the results from these experiments supported the hypothesis that illusory learning occurs with unfamiliar tunes; however, familiar tunes did not show any dissociation between metamemory judgments and memory performance. This was an unexpected finding, particularly due to the fact that previous literature exploring the illusory learning phenomenon used familiar words (Rhodes & Castel, 2008; Rhodes & Castel, 2009), which we thought would be more analogous to familiar tunes. It appears as though people use ease of processing to help determine the likelihood an item will be remembered at a later time for written and spoken words as well as unfamiliar tunes, but another, more accurate strategy is used to make metamemory judgments for familiar tunes. Conversely, metacognitive judgments were found to be more accurate for unfamiliar texts as compared to familiar texts (Johnson & Halpern, 1999). In general, this finding suggests that a dissociation exists between familiar and unfamiliar tunes as well as between words and music.

In addition, we found that whereas participants with both low and high levels of musical experience predicted better memory performance for easier to process unfamiliar items as compared to harder to process unfamiliar items, this prediction was only accurate among participants with low levels of musical experience; more trained people showed no difference in memory performance for easier and harder to process items. In contrast, Experiment 3 revealed no effect of musical background on ease of processing for either JOL judgments or memory performance with familiar items. These results raise the possibility that people are generally well calibrated on a gross level to their future memory performance for familiar tunes, but are unaware of how continued musical experience affects memory performance for unfamiliar tunes.

Furthermore, we found that metamemory accuracy on an item-by-item basis was far better for familiar tunes as compared to unfamiliar tunes. While it appears as though metamemory for nonsense words, which are more analogous to unfamiliar tunes, has not been explored, an experiment comparing metamemory judgments for verb-noun word pairs (i.e. bend knee) and for novel verb-noun word pairs (i.e. smell knee) found that metamemory judgments were most accurate for novel word pairs (McDonald-Miszczak, Hubley, & Hultsch, 1996). This seems to suggest a difference between words and music where item-by-item metamemory for unfamiliar word pairs and familiar music is better than for familiar word pairs and unfamiliar music. However, it is important to note that novel word pairs and unfamiliar music may not be comparable; novel word pairs are composed of two words with meaning that are put together in a novel way whereas unfamiliar music has no meaning.

Although we did not design the study to investigate a memory difference between fast and slow tunes, Dowling, Bartlett, Halpern, and Andrews (2008) found similar results in an experiment where participants first heard a list of tunes that were played in slow, intermediate, and fast tempi. Immediately following each tune, subjects heard another tune that was either the same as the original tune or had two pitches changed. Participants had to indicate whether this tune was the same or different as compared to the original tune. They found that recognition memory performance for familiar tunes was best for tunes played in an intermediate tempo and significantly worse for tunes played in slow tempo followed by tunes played in fast tempo. Although there are procedural differences between the experiments, this research appears to validate our finding that slow tunes were better remembered than fast tunes. An additional discovery from our experiment suggests that participants were also unaware of the memory difference because metamemory judgments were statistically the same for fast and slow tunes. This appears to be another example of a dissociation between metamemory and memory, although it was a reverse dissociation from what we originally predicted.

Procedurally, we found that grouping the study and test phases for the volume and timbre blocks for Experiment 2 improved memory performance from Experiment 1. In conjunction with changing the JOL scale and instructions, we found the illusory learning effect with unfamiliar tunes in Experiment 2 when we did not see this effect in Experiment 1. Overall, it appears that slightly boosting memory performance and maximizing the range of JOL values used allowed us to more accurately assess the relationship between metamemory and memory.

In real life, it appears that while metamemory for familiar tunes is not perfect, it is accurate enough that it can be considered a reliable measure of future recognition memory performance. However, caution should be taken when monitoring memory for unfamiliar tunes; metamemory judgments appear to be influenced by ease of processing, suggesting that people are more likely to overestimate their memory for a new melody if it is played, for example, at a louder volume or in a more familiar instrument. This information could be relevant for musicians who are learning new pieces to perform.

Collectively, these results suggest that people do not use one process to make metamemory judgments for all types of information; instead, it appears as though different strategies are used for different types of information, such as familiar versus unfamiliar tunes. On a broader scale, this raises the possibility that the brain treats different types of information distinctly for multiple cognitive abilities. For example, researchers have begun to compare the brain areas involved in processing music and language. An experiment by Schmithorst (2005) compared fMRI activation for participants passively listening to music to areas previously shown to be activated in language. He found that while there are overlaps in areas of activation, such as in areas believed to process syntax, there also appears to be brain activation that is unique to music processing. Brain activation for metamemory was studied by Chua, Schacter, and Sperling (2008), who found that areas of the prefrontal and parietal lobes were activated in metamemory tasks, but not in control, non-metamemory tasks. Although neural correlates of metamemory for different types of information have not yet been compared, future research directions include exploring these potential differences and eventually determining how we process and monitor all forms of information.

References

Bartlett, J.C., Halpern, A.R., & Dowling, W.J. (1995). Recognition of familiar and unfamiliar melodies in normal aging and Alzheimer's disease. *Memory & Cognition*, *23*, 531-546.

Bunnell, J.K., Baken, D.M., & Richards-Ward, L.A. (1999). The effect of age on metamemory for working memory. *New Zealand Journal of Psychology*, *28*, 23-29.

Chandler, M.J., Lacritz, L.H. , Cicerello, A.R., Chapman, S.B., Honig, L.S., Weiner, M.F., & Cullum, C.M. (2004). Three-word recall in normal aging. *Journal of Clinical and Experimental Neuropsychology*, *26*, 1128-1133.

Chua, E.F., Schacter, D.L., & Sperling, R.A. (2008). Neural correlates of metamemory: a comparison of feeling-of-knowing and retrospective confidence judgments. *Journal of Cognitive Neuroscience*, *21*, 1751-1765.

Dowling, W.J., Bartlett, J.C., Halpern, A.R., & Andrews, M.W. (2008). Melody recognition at fast and slow tempos: effects of age, experience, and familiarity. *Perception & Psychophysics*, *70*, 496-502.

Dunlosky, J., Baker, J.M.C., Rawson, K.A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*, 390-400.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: SAGE Publications, Inc.

Fleischman, D.A., Wilson, R.S., Gabrieli, J.D.E., Bienias, J.L., & Bennett, D.A. (2004). A longitudinal study of implicit and explicit memory in old persons. Psychology and Aging, 19, 617-625.

Halpern, A.R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology*, *61*, 1371-1384.

Halpern, A.R., & Bartlett, J.C. (2010). Memory for melodies. In M.R. Jones, A.N. Popper, R.R. Fay (Eds.), *Music Perception*. New York, NY: Springer-Verlag.

Hertzog, C., Dixon, R.A., & Hultsch, D.F. (1990). Relationships between metamemory, memory predictions, and memory task performance in adults. *Psychology and Aging*, *5*, 215-227.

Jackendoff, R. (2009). Parallels and nonparallels between language and music. *Music Perception: An Interdisciplinary Journal*, *16*, 195-204.

Johnson, S.K., & Halpern, A.R. (1999). Prediction accuracy of young and middle-aged adults in memory for familiar and unfamiliar texts. *The American Journal of Psychology*, *112*, 235-257.

Justice, E.M., & Weaver-McDougall, R.G. (1989). Adults' knowledge about memory: awareness and use of memory strategies across tasks. *Journal of Educational Psychology*, *81*, 214-219.

Korenman, L.M., & Peynircioglu, Z.F. (2004). The role of familiarity in episodic memory and metamemory for music. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 917-922.

Lyons, L. (2003, May 20). *Americans want music students to play on*. Retrieved from http://www.gallup.com/poll/8434/Americans-Want-Music-Students-Play.aspx.

Maki, R.H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 116-126.

Maki, R.H., Willmon, C., & Pietan, A. (2009). Basis of metamemory judgments for text with multiple-choice, essay, and recall tests. *Applied Cognitive Psychology*, *23*, 204-222.

McDonald-Miszczak, L., Hubley, A.M., & Hultsch, D.F. (1996). Age differences in recall and predicting recall of action events and words. *The Journals of Gerontology*, *51B*, P81-P90.

McGillivray, S., & Castel, A.D. (2010). Memory for age-face associations in younger and older adults: the role of generation and schematic support. *Psychology and Aging*, *25*, 822-832.

Peynircioglu, Z.F., Rabinovitz, B.E., & Thompson, J.L.W. (2008). Memory and metamemory for songs: the relative effectiveness of titles, lyrics, and melodies as cues for each other . *Psychology of Music*, *36*, 47-61.

Peynircioglu, Z.F., Tekcan, A.I., Wagner, J.L., Baxter, T.L., & Shaffer, S.D. (1998). Name or hum that tune: feeling of knowing for music. *Memory & Cognition*, *26*, 1131-1137.

Rawson, K.A., & Dunlosky, j. (2002). Are performance predictions for text based on ease of processing?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 69-80.

Rhodes, M.G., & Castel, A.D. (2008). Memory predictions are influenced by perceptual information: evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615-625.

Rhodes, M.G., & Castel, A.D. (2009). Metacognitive illusions for auditory information: effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*, 550-554.

Schmithorst, V.J. (2005). Separate cortical networks involved in music perception: preliminary functional mri evidence for modularity of music processing. *NeuroImage*, *25*, 444-451.

Schulkind, M.D., Hennis, L.K., & Rubin, D.C. (1999). Music, emotion, and autobiographical memory: they're playing your song. *Memory & Cognition*, *27*, 948-955.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for owrds. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381-391.

Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, *54*, 298-303.

Voss, J.L., & Paller, K.A. (2008). Brain substrates of implicit and explicit memory: the importance of concurrently acquired neural signals of both memory types. *Neuropsychologia*, *46*, 3021-3029.

Weaver, C.A., & Bryant, D.S. (1995). Monitoring of comprehension: the role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*, 12-22.

Yeates, K.O., & Enrile, B.G. (2005). Implicit and explicit memory in children with congenital and acquired brain disorder. *Neuropsychology*, *19*, 618-628.

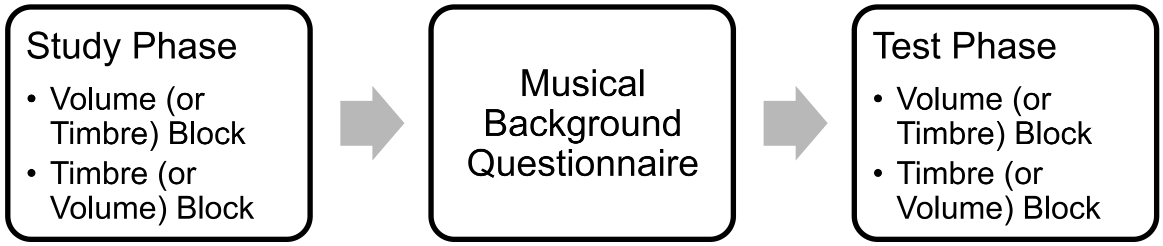


Figure 1. Schematic of Experiment 1 procedure including study phase, questionnaire, and test phase.

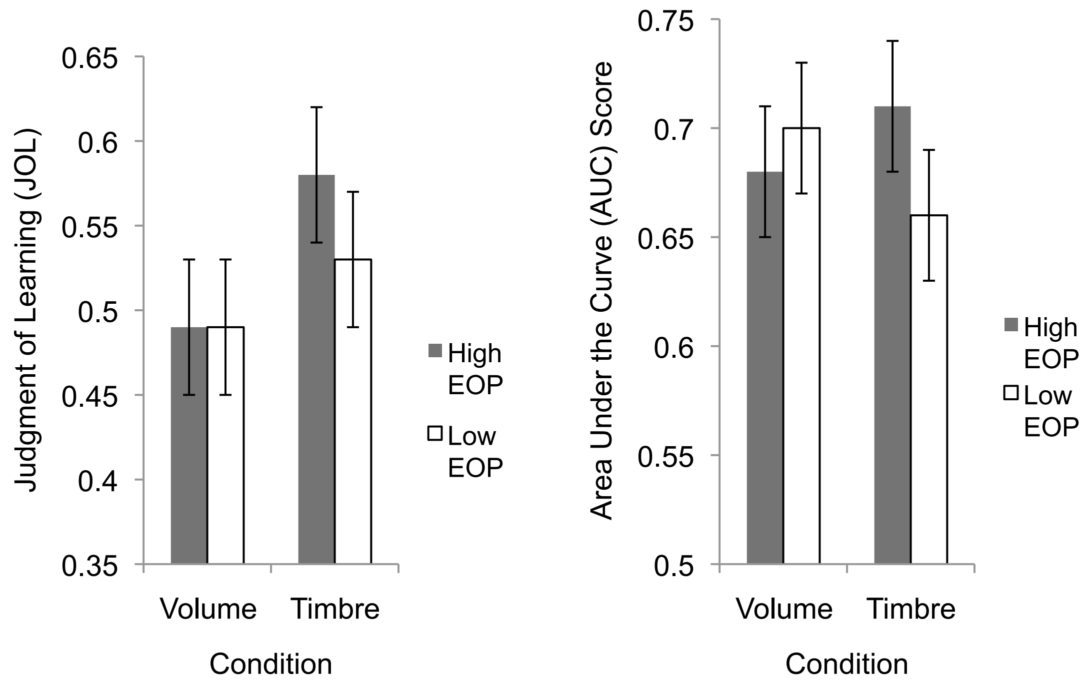


Figure 2. Mean JOL and AUC Scores (± SE) for the volume and timbre conditions in Experiment 1. EOP = ease of processing level.

Figure 3. Schematic of Experiment 2 procedure including volume (or timbre) block, questionnaire, and timbre (or volume) block.

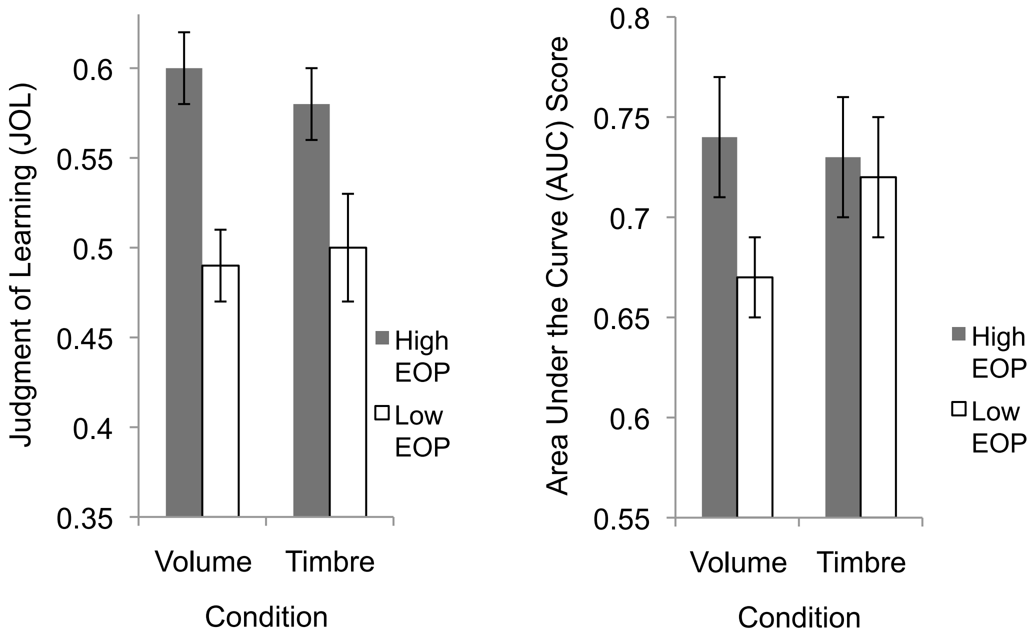


Figure 4. Mean JOL and AUC Scores (± SE) for the volume and timbre conditions in Experiment 2. EOP = ease of processing level.

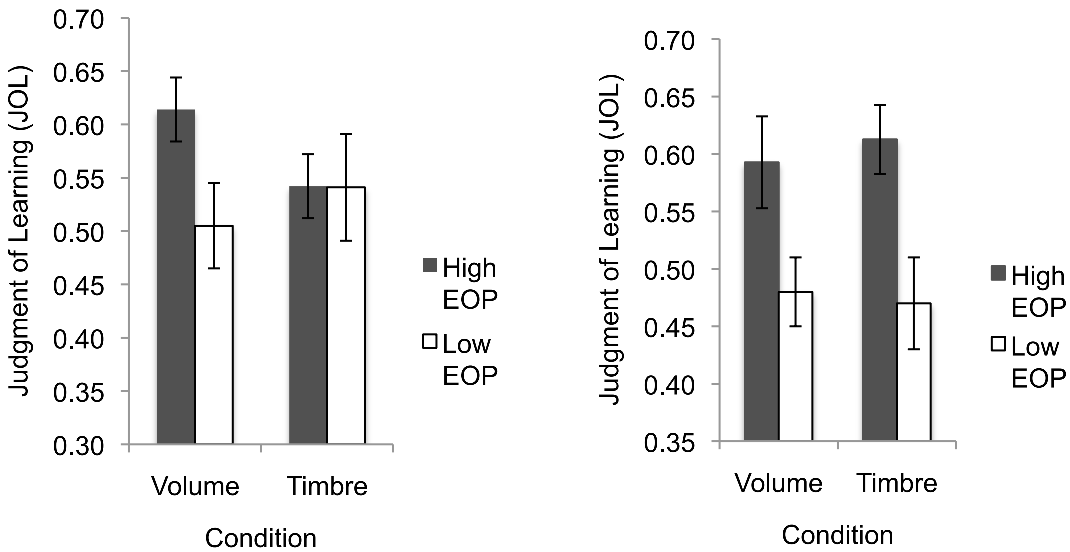


Figure 5. Mean JOLs (± SE) for the volume and timbre conditions for participants with low (left) and high (right) levels of musical experience. EOP = ease of processing level.

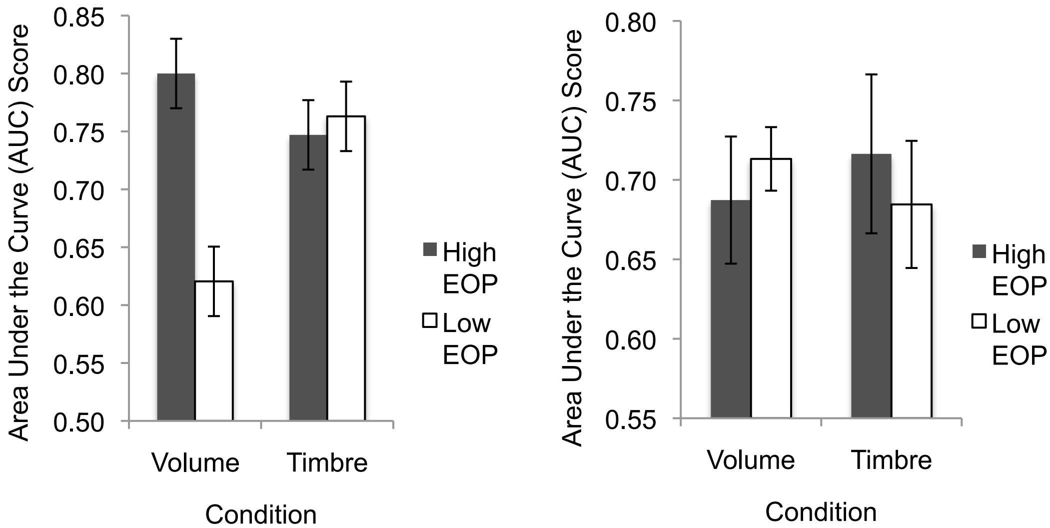
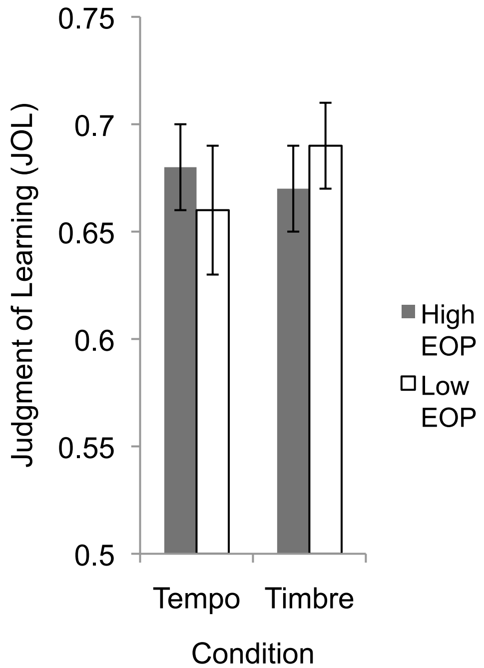


Figure 6. Mean AUC Scores (± SE) for the volume and timbre conditions for participants with low (left) and high (right) levels of musical experience. EOP = ease of processing level.

Figure 7. Schematic of Experiment 3 procedure including tempo (or timbre) block, questionnaire, and timbre (or tempo) block.



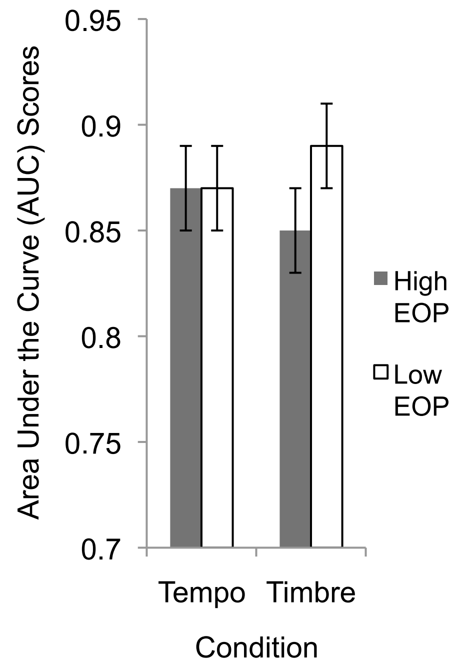


Figure 8. Mean JOL and AUC Scores (± SE) for the tempo and timbre conditions in Experiment 3. EOP = ease of processing level.