2011

Levels-of-Processing Effects on "Remember" Responses in Recognition For Familiar and Unfamiliar Tunes

Esra Mungan

Zehra F. Peynircioğlu

Andrea R. Halpern
Bucknell University, ahalpern@bucknell.edu

Levels-of-Processing Effects on "Remember" Responses in Recognition for Familiar and Unfamiliar Tunes

ESRA MUNGAN
Bogazici University

ZEHRA F. PEYNİRCİOĞLU
American University

ANDREA R. HALPERN
Bucknell University

We investigated the effect of level-of-processing manipulations on "remember" and "know" responses in episodic melody recognition (Experiments 1 and 2) and how this effect is modulated by item familiarity (Experiment 2). In Experiment 1, participants performed 2 conceptual and 2 perceptual orienting tasks while listening to familiar melodies: judging the mood, continuing the tune, tracing the pitch contour, and counting long notes. The conceptual mood task led to higher d' rates for "remember" but not "know" responses. In Experiment 2, participants either judged the mood or counted long notes of tunes with high and low familiarity. A level-of-processing effect emerged again in participants' "remember" d' rates regardless of melody familiarity. Results are discussed within the distinctive processing framework.

The level-of-processing (LOP) effect (Craik & Lockhart, 1972) is one of the most researched effects in memory literature. Processing items deeply, or semantically, leads to better memory performance than processing them at a shallow level or with respect to their physical or perceptual properties (Craik & Tulving, 1975; Craik & Watkins, 1973; Friedman & Bourne, 1976; Jacoby, Craik, & Begg, 1979). Interestingly, this memory advantage for deeply processed items over shallowly processed ones occurs regardless of whether a memory test is expected (Hyde, 1973; Hyde & Jenkins, 1973), suggesting that semantic processing seems to be sufficient for learning to occur (cf. Craik, 2002).

The LOP effect in episodic recognition memory has been shown for both verbal and various nonverbal materials such as faces (Bower & Karlin, 1974; Warrington & Ackroyd, 1975), pictures (Gardiner, Gregg, & Karayianni, 2006; Marks, 1991), and voices (Church & Schacter, 1994; Schacter & Church, 1992). However, for musical materials, only one study has shown an LOP effect (Peretz, Gaudreau, & Bonnel, 1998), and a few have failed to show such an effect (Halpern & Müllensiefen, 2008; Warker & Halpern, 2005).

In their 1998 study, Peretz, Gaudreau, and Bonnel presented nonmusicians with a mixed list of familiar and unfamiliar tunes to be processed either

by rating their familiarity level (familiarity encoding task) or by noting their timbre (instrument encoding task). The former task was assumed to be a deep processing task in that it focused the participants' attention on the more abstract, conceptual aspects of the material, whereas the latter was assumed to be a shallow processing task in that it focused participants' attention on the more physical aspects of the material. The findings revealed an LOP effect in the expected direction but only with familiar melodies. In the Warker and Halpern (2005) study, participants had to rate either the pleasantness (semantic task) or the rhythmic regularity (nonsemantic task) of a set of obscure folk tunes. Findings did not reveal an LOP effect, thus replicating Peretz et al.'s (1998) finding for unfamiliar tunes. Halpern and Müllensiefen (2008) also failed to find an LOP effect for unfamiliar tunes. Thus, unfamiliar music appears to be the only type of material in the LOP memory literature that consistently does not show this effect. Because LOP effects occur with both high- and low-frequency words (Duchek & Neely, 1989; Mandler, Goodman, & Wilkes-Gibbs, 1982), legal and illegal nonwords (Bowers, 1994), and famous and nonfamous faces (a closer analogy to familiar and unfamiliar tunes; Wig, Miller, Kingstone, & Kelley, 2004) the goal of our research was to address why LOP does not seem to influence music, especially unfamiliar music, the way it does other materials.

In the present study, we took finer measures of recognition memory by using "remember" and "know" judgments. Tulving (1985) differentiated between the subjective state of being able to consciously retrieve the moment when a critical item was encountered ("remember") and the subjective state of retrieving information without recollecting details about when it was first encountered ("know"). LOP manipulations typically affect "remember" (R) but not "know" (K) responses (Gardiner, 1988; Gregg & Gardiner, 1994; Yonelinas, 2001), presumably because the remembering experience is more closely linked to episodic memory processes and therefore is sensitive to distinctive processing, whereas the knowing experience is more closely linked to semantic memory processes and therefore is sensitive to perceptual fluency-related factors (cf. Rajaram, 1998; Rajaram & Geraci, 2000). This mechanism might serve at least as a proximate explanation for the LOP effect in episodic memory

because if conceptual processing increases recognition performance in R- but not K-type responses, it must be the R responses that create the general LOP effect in memory performance. Therefore, using R and K measures in LOP studies with music may help us understand why the effect seemingly occurs with familiar but not with unfamiliar music. That is, perhaps deep processing tasks are able to create sufficient distinctiveness to increase R responses with familiar music but not with unfamiliar music.

In Experiment 1, participants listened to familiar tunes while engaged in different orienting tasks (OTs). During the ensuing recognition test, "remember" and "know" measures were taken in addition to "yes" and "no" responses. OTs used in LOP research can be broadly classified as conceptual processing tasks (e.g., giving pleasantness ratings or judging familiarity of items) and perceptual processing tasks (e.g., counting the number of vowels or finding rhyming words of items), a distinction that has come to be preferred over the earlier, somewhat problematic distinction between deep and shallow processing tasks. Most research uses one of each type of task to demonstrate LOP effects; however, because the conceptual–perceptual distinction may be more complicated when it comes to music (cf. Halpern & Müllensiefen, 2008), we decided to use two conceptual and two perceptual OTs. The four OTs were judging what kind of mood the melody conveys (Conceptual Task 1), continuing the melody by imagining the next two to three notes (Conceptual Task 2), tracing the melody contour (Perceptual Task 1), and counting the number of long notes (Perceptual Task 2). The latter two tasks were expected to involve more data-driven processing, whereas continuing a melody and judging its mood were expected to involve more conceptual processing. For instance, Meyer (1956) wrote about musical stimuli as activating expectations and tensions about their continuation that, depending on whether fulfilled or not, form its affective content. Given that conceptual processing is defined as processing in which one engages in interpretive processes and thus goes beyond what is immediately present, our first two tasks should qualify as such (cf. Palmer, 1997; Repp, 1992). However, because none of these OTs have been researched extensively within the LOP framework in music recognition, we wanted to also look at the effect of each task separately.

Thus, we hypothesized that the conceptual tasks compared to the perceptual tasks would lead to more distinctive processing and thus increase the recognition sensitivity in R responses in the subsequent recognition test. We were also curious as to whether our LOP effect on recognition sensitivity in R responses would be strong enough to create one in overall recognition sensitivity measures.

In Experiment 2, we explored how familiarity might modulate the effects of LOP on recognition sensitivity of R and K responses in music. We suspected that the failure of previous research to find LOP effects with unfamiliar music might have resulted from the ineffectiveness of LOP manipulations in increasing recognition sensitivity in R responses to begin with. In other words, unfamiliar music, unlike unfamiliar faces (Mäntylä, 1997), may simply be a type of material that is not amenable to the formation of distinctive episodes.

To summarize, the goals of the present study were twofold. First, we examined the elusive nature of LOP effects in music by using R and K responses to serve as finer measures of episodic recognition memory and a greater range of potentially more pure conceptual and perceptual tasks compared with the tasks used so far with music (Experiment 1). Second, we examined how melody familiarity might modulate such effects (Experiment 2). We confined our investigation to nonmusicians both because we were interested in the memory experiences of ordinary listeners and also because pilot work indicated that musical experience did not interact with our other measures of interest.

EXPERIMENT 1

METHOD

Participants

Thirty-two nonmusicians, who were undergraduate students of Boğaziçi University and received extra credit in psychology courses, served as participants. All had had less than 1 year of musical training and could not read music. To be eligible, participants had to pass a transposition test, in which they had to differentiate between a minor and a major triad interval. This test was necessary because our recognition test contained lures that were often very similar to the targets in terms of pitch sequence. Six participants who did not pass the transposition test were excluded and replaced to reach a sample size of 32 participants.

Materials and Design
The materials were excerpts from classical music, soundtracks, dance music, and songs. They had either no lyrics (e.g., an excerpt from the soundtrack of the movie *Pulp Fiction* or from *A Little Nightmusic* by Mozart) or their lyrics were not commonly known (e.g., an excerpt from *Carmen* by Bizet).

Through extensive piloting of 173 melodies on a group of 51 participants, a group of highly familiar excerpts was selected. These were 48 excerpts from 39 different pieces. The excerpts were yoked so that for each target excerpt, there was a similar-sounding lure excerpt in the recognition test. In 10 of the 24 target–lure pairs, both excerpts were from the same piece; in the remaining 14 pairs, they were from different pieces.[1] Because some of the lures were also from the same pieces as the targets, knowing or rehearsing the name of the piece would be eliminated as a factor in the participants' performance. The excerpts averaged about 11.3 s, with a range from 7 to 21 s, and the lures and targets were of similar duration on average. Two study lists (A and B), each with 24 excerpts, and one common recognition test with all 48 excerpts were recorded for presentation. For one group of participants, the A list comprised the targets and the B list the lures and vice versa for the other group. All excerpts were played on a Yamaha S03 keyboard and recorded as MIDI files in piano timbre. The MIDI files were then transformed into WAV files and recorded on two CDs to serve as Study List A and Study List B. The 48-item recognition test CD was constructed by randomly shuffling the 24 A and 24 B excerpts. The interstimulus intervals between excerpts in both study and test lists were 4 s.

Four OTs (counting, contour tracing, continuation, and mood judgment) were counterbalanced across melodies in blocks of six. Thus, four different versions of study lists were constructed so that each block of excerpts would be studied with each OT equally often across four subgroups of participants. Each study list sheet was numbered 1 through 24, to refer to the melodies that the participants were about to hear. Each number was followed by the particular OT indicator. Participants received a booklet of three pages during the listening phase, including one practice page and two study pages.

A practice CD was prepared to familiarize participants with the different OTs. The CD contained two

practice trials for each of the four OTs. None of the practice melodies were from the experimental sets.

The recognition test sheet consisted of 48 lines spread over four columns numbered 1 through 48, with a "Yes" and "No" next to each number. The remaining handouts were a musical background questionnaire and a familiarity rating sheet on which to give familiarity ratings to all 48 melodies of the recognition test.

The overall design was a two-factor 4 (OT: mood, continuation, contour, counting) × 3 (response type: remember, know, guess) within-participant design. The dependent measure was recognition sensitivity (d').

Procedure

Before the experiment, participants went through a two-trial transposition test recorded on CD. First, they heard an example of a correct transposition, with "Twinkle Twinkle Little Star" played starting on C (CCGGAAG) and then from F (FFCCDDC) and an incorrect transposition from F (FFCCD♭D♭C). They were then told that they would hear two different trials of two 3-note sequences and that on one trial the two sequences would be alike (i.e., correctly transposed), and on the other they would be different from each other (i.e., incorrectly transposed). They then heard Trial 1 as CED–FA♭G (incorrect transposition) and Trial 2 as CED–FAG (correct transposition). For each trial they had to indicate whether they thought the second sequence sounded the same as or different from the first. After participants were reminded that in one case they were the same and in the other different, the two trials were repeated.

Participants were told that they were going to listen to various melodies on which they would have to perform certain perception tasks. To better follow the instructions, participants were given their OT booklet with the practice page on top. They were told that for some of the melodies (counting task) they should try to count the number of long notes and then write down that number in the slot provided next to the trial number ("1. ___ long notes"). Instructions emphasized that this task was also a purely subjective task because long is a relative term. Therefore, they should set up in advance their own subjective criterion of a "long note" and then count the notes that fit their criterion as the melody went along. In the contour task they were asked to try to trace the melodic contour (increases and decreases in pitch) of the melody. We explained that the symbol followed by the horizontal line next to the trial number repre-

sented the pitch height of the first note in the melody ("7. ▫ _____"). Listeners were asked to put their pencils onto the bar before the melody started and to trace, as best as they could, the changes in pitch height as the melody went along. In the mood task, they had to indicate the type of mood that the melody conveyed by circling any of the given options ("sad," "happy," "angry," "depressive," "dramatic"), and if they found none of the given options adequate, they should circle the "other: _____" option and write down a word that for them best described the mood of the melody. Finally, in the continuation task participants were asked to try to think of how the melody could continue. In particular, they had to indicate what the next two or three notes could be to follow the sequence. They were told that the symbol that preceded the horizontal line represented the pitch height of the last note of the melody ("18. ▫ _____") and that it was there as a rough guide for them to gauge the pitch height of the next two notes they would imagine (if the melody was familiar to them they were told they could use the actual continuation of the melody). Instructions emphasized that they should use the line just as a reference to mark approximate pitch relations. Other than the contour tracing task, there were no right and wrong answers; therefore, we scored only that one task.

After these instructions, participants completed practice trials for the different tasks until they felt at ease with the tasks before going on to the actual study. At this stage, no mention was made of any memory test to follow. After the listening phase was over, participants were surprised with a recognition test. They were handed answer sheets and told that their memories would be tested for the 24 melodies they had just heard. For each melody they would have to indicate whether it was one of the 24 melodies they had listened to, by circling "yes" or "no." They were then given the standard instructions for "remember," "know," and "guess" responses. They were told that if they were sure they recognized the tune and remembered at least one thing about the preceding episode during which they had encountered the tune (e.g., they may remember which task they performed while listening to it or remember that when hearing that tune a personal memory came to their mind) they were to mark an "R" next to their "yes" response. If they were sure they had heard the tune in the preceding phase but could not come up with any specific detail about that instance of encounter, they were to mark a "K" next to their "yes" response. If they were not sure about having heard that tune

during the first phase but felt as if they had, they were to mark a "G" next to their "yes" response. "Guess" responses were taken in order to prevent contamination of K responses with guessing. It was pointed out that they should listen carefully before making their decision because there might be cases in which the lure melody might be very similar to the one they had heard, differing from the latter in only a few notes.

Participants then completed a questionnaire about their musical background, their approximate amount of weekly exposure to music, and their musical preferences. Finally, participants listened to all 48 excerpts of the recognition test once again and rated each one in terms of familiarity. A piece so familiar that they could easily have sung it themselves and also its continuation (regardless of whether they could come up with the name of the piece) was rated 10 (*highly familiar*). Melodies they felt they had never heard before were rated 1 (*very unfamiliar*). Intermediate numbers indicated intermediate degrees of familiarity. Participants were tested individually or in groups of up to six people. Each session lasted about 35 min.

RESULTS

Participants' familiarity ratings for the 48 melodies used in Experiment 1 were in line with the pilot familiarity rating results; on a 10-point scale they averaged a highly familiar rating of 8.5 ($SD = 1.9$), and there were no differences between the two presentation lists.

Recognition Sensitivity Analyses
We did two sets of analyses with respect to recognition sensitivity. In the first one, we simply looked at participants' overall d' scores irrespective of their response types, whereas in the second analysis we included their R, K, and G judgments.

For the first set of analyses, recognition data were converted into d' scores, using each participant's OT-specific hit rates and their OT-nonspecific false alarm rates. A one-factor ANOVA on these d' scores revealed a significant main effect for OT, $F(3, 93) = 2.88$, $MSE = 0.99$; $p < .05$, partial $\eta^2 = .09$. A subsequent paired t test comparing the two conceptual OTs with the two perceptual OTs was significant, with the former yielding higher d' rates ($M = 1.57, SD = 0.64$) than the latter ($M = 1.18, SD = 0.58$), $t(31) = 2.72, p < .05$.

For the second set of analyses, recognition data were converted into d' scores, using each participant's OT-specific R, K, and G hit rates and their condition-nonspecific R, K, and G false alarm rates, thus obtaining R, K, and G d' scores across four OT conditions. As can be seen in Figure 1, this led to lower d' values, because a total hit score was split into its respective R, K, and G components.

A 4 (OT: mood, continuation, contour, counting) × 3 (response type: R, K, G) repeated-measures ANOVA on d' rates yielded no main effect for OT, $F(3, 93) = 1.28$, $MSE = 0.37$; $p > .10$, partial $\eta^2 = .04$.[2] There was also no main effect for response type, $F(2, 62) = 2.27$, $MSE = 2.01$, $p > .10$, partial $\eta^2 = .07$. However, there was a significant interaction between OT and response type, $F(6, 186) = 2.66$, $MSE = 0.81$, $p > .05$, partial $\eta^2 = .08$. When we look at Figure 1, we see that both R and G d' scores change across OT conditions, and the mood task appears to create the
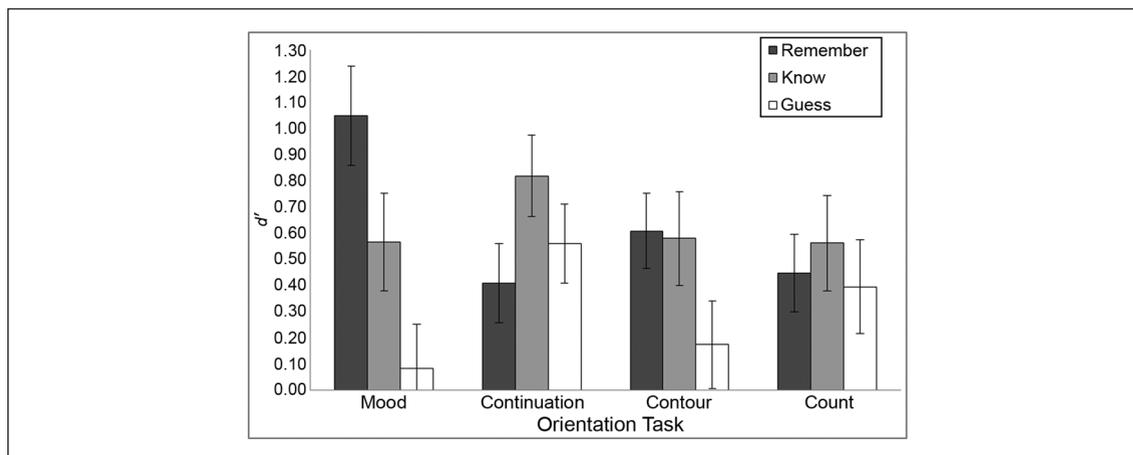

FIGURE 1. d' values with *SE*s as a function of orienting task and response type, Experiment 1

largest discrepancy. G d' scores actually seem to mimic R d' scores but in an opposite way. Compared with R and G d' scores, K d' scores appear to be minimally influenced by OT condition. Separate one-factor ANOVAs performed on R, K, and G d' scores showed significant or nearly significant OT main effects for R and G d' scores, $F(3, 93) = 4.89$, $MSE = 0.56$, $p < .01$, partial $\eta^2 = .17$ and $F(3, 93) = 2.56$, $MSE = 0.59$, $p = .06$, partial $\eta^2 = .08$, respectively, but not for K d' scores, $F(3, 93) < 1$, $p > .10$. Bonferroni post hoc tests of comparison with a p value set at .05 revealed that the mood task produced significantly higher R d' scores than any of the other tasks.

Recognition Bias Analyses

For sake of comparison, we analyzed participants' recognition bias tendencies as measured by c. Negative c scores signify a "yes" bias, whereas positive ones signify a "no" bias (Stanislaw & Todorov, 1999).

In c scores using OT-specific hit rates and OT-nonspecific false alarm rates while disregarding response type, we found a similar OT main effect in response bias, $F(3, 93) = 2.88$, $MSE = 0.25$, $p < .05$, partial $\eta^2 = .09$. A subsequent paired t test comparing the two conceptual OTs with the two perceptual OTs revealed a stronger "no" bias for conceptually processed tunes ($M = .29$, $SD = .49$) than for perceptually processed ones ($M = .10$, $SD = .40$), $t(31) = 2.72$, $p < .05$.

The c scores using OT-specific R, K, and G hit rates and OT-nonspecific R, K, and G false alarm rates in a 4 (OT) $\times$ 3 (response type) ANOVA revealed neither a main effect of OT nor a main effect of response type (ps > .10). However, there was a significant interaction between OT and response type, $F(6, 186) = 2.66$, $MSE = 0.20$, $p < .05$, partial $\eta^2 = .08$. Subsequent one-factor ANOVAs revealed no OT effect on K, a marginal OT effect on G response biases ($p = .06$), but a significant OT effect on R response biases, $F(3, 93) = 4.89$, $MSE = 0.14$, $p < .05$, partial $\eta^2 = .09$, with Bonferroni post hoc tests showing that the mood task produced a lesser "yes" bias in R responses ($M = -0.85$) than the continuation and counting tasks ($M = -1.17$ and $M = -1.15$, respectively).

DISCUSSION

In our first study we found that, when grouped together, the more conceptually driven tasks led to higher recognition sensitivity rates than the more perceptually driven tasks. This finding is in line with Peretz et al.'s (1998) LOP finding with familiar tunes.

However, the primary purpose of Experiment 1 was to explore LOP effects on different types of recognition responses. We did indeed find an LOP effect on recognition sensitivity of R but not K responses; interestingly, the mood task seemed to stand out from the rest by producing the highest R d' rates. We did not observe a similar effect with our other conceptual task, the continuation task. Two possible explanations come to mind. One is that the continuation task, even though thought of as a conceptual task (i.e., a task that engaged the listener in interpretive, semantic-like processing) was not in fact conceptual. This possibility is further corroborated by the fact that the continuation task was the only one of the four tasks that produced—though nonsignificantly—higher K than R d' rates (.82 vs. .41). Another possibility is that whether or not a task is conceptual may not predict LOP; instead, the degree of distinctive processing it involves may be the critical aspect. In that sense, our mood task, which involved a cognitive appraisal of the mood experienced when hearing the melody, might have been both conceptual and distinctive, whereas our continuation task might have been conceptual but not distinctive. The finding that the mood task was uniquely successful in increasing recognition sensitivity in R responses will be discussed further later on.

The absence of an LOP effect on K d' scores was in line with earlier findings using nonmusical materials (e.g., Gardiner, 1988; Gregg & Gardiner, 1994; Yonelinas, 2001) because K responses are believed to be more sensitive to perceptual fluency-related factors such as same versus different modality between study and test rather than distinctive processing-related processes (cf. Rajaram, 1998; Rajaram & Geraci, 2000). An interesting finding was the "reverse" LOP trend in G d' scores, showing the opposite effect that was observed in R d' scores. Perhaps an OT that decreases distinctiveness increases the sensitivity of G type responses, and vice versa for an OT that increases distinctiveness.

Interestingly, participants' bias scores revealed an overall "no" bias when "old" responses were combined and an overall "yes" bias when these were separated into R, K, and G judgments. In either case, compared with the other tasks, the mood task stood out in pro-

ducing the highest "no" bias when response type was ignored ($M = 0.47$ as opposed to 0.35, 0.13, and 0.21 for continuation, contour, and counting, respectively) and the lowest "yes" bias in R ($M = -0.85$ as opposed to -1.17, -1.07, and -1.15 for continuation, contour, and counting). In other words, participants appeared to be more careful to say "yes" for melodies processed with the mood task and be more prone to saying "no" when giving R judgments at recognition.

EXPERIMENT 2

Because LOP effects have so far not been obtained with unfamiliar musical stimuli (Peretz et al., 1998; Warker & Halpern, 2005), Experiment 2 tested whether and how the LOP effect on recognition sensitivity in R responses would change depending on melody familiarity. By using two different levels of familiarity (high and low) and two OTs (mood judgment and counting long notes, the two tasks that produced the most extreme recognition sensitivity levels in R responses in Experiment 1), we explored whether the LOP effect on recognition sensitivity in R responses was limited to familiar musical materials or whether earlier null findings with unfamiliar music were due to the specific OTs (Halpern & Müllensiefen, 2008).

Mäntylä (1997) found that distinctive processing of unfamiliar faces increased the number of correct R responses. We were curious whether we, too, would obtain such a benefit in recognition sensitivity in R responses as a result of conceptual processing even for low-familiarity melodies—even if that increase was not sufficient for an overall LOP effect—or whether unfamiliar music would be immune to any LOP effect, including one in R sensitivity. As is usually the case, we did not expect the OT manipulation to have an effect on recognition sensitivity of K responses. In addition, because melody familiarity was varied between lists[3] it was necessary to switch to an intentional memory procedure, with instructions that informed participants in advance that a memory test would follow.

METHOD

Participants

Forty-eight nonmusicians, who were undergraduate students of Boğaziçi University and received extra credit in psychology courses, served as participants. All participants had had less than 1 year of musical training and could not read music. Nine participants who did not pass the transposition test were excluded and replaced to reach a sample size of 48 participants.

Materials, Design, and Procedure

The materials were the same as in Experiment 1 except that in addition to the 24-melody study list of high-familiarity melodies used in Experiment 1, an additional 24-melody list of low-familiarity melodies was constructed. The low-familiarity list also had two versions for target–lure counterbalancing purposes. All low-familiarity melodies were chosen from the same pool of 173 melodies that had been piloted for Experiment 1, so they had roughly the same musical properties (all keyboard-recorded, single-line music excerpts, mostly from instrumental pieces, similar tune lengths). The high-familiarity melodies were those that had an average familiarity rating between 9 and 10 ($M = 9.3$, $SD = 1.7$) on a 10-point scale. Low-familiarity melodies were those that had an average familiarity rating between 1 and 4 ($M = 3.0$, $SD = 2.7$). There were two recognition tests, one that consisted of 48 high-familiarity melodies and one that consisted of 48 low-familiarity melodies, with 24 targets and 24 lures in each case. For half of the participants, List A melodies served as the targets and List B melodies as the lures, and vice versa for the other half.

The study booklets were the same as in Experiment 1, with the exception that they were grouped according to the eight counterbalancing conditions, list type (List A vs. List B), familiarity order (high familiarity first vs. low familiarity first), and OT task order (counting first vs. mood first), and the recognition test sheets required R, K, or G responses for each "yes" response. The overall design was a 2 (OT: mood judgment vs. counting long notes) × 2 (melody familiarity: high vs. low) × 3 (response type: R, K, G) within-participant design with d' as the main dependent measure. The procedure was the same as in Experiment 1 except that participants went through two study–test phases, one for high-familiarity melodies and one for low-familiarity melodies. The fact that each study list consisted of melodies of a different level of familiarity was not explicitly mentioned. Participants were tested in groups of up to four. The experiment was conducted in two sessions. The first session lasted for about 55 min and included the two study–test phases, which were administered successively without delay. In the second session, participants listened to all 96 melodies and rated their familiarity. The second session took about 15–20 min.

RESULTS

The subjective familiarity ratings were similar to the ratings obtained during pilot testing; the high-familiarity melodies were given an average rating of 9.1 (*SD* = 2.1) and the low-familiarity melodies 3.9 (*SD* = 2.5). There were also no significant differences between the subjective familiarity ratings for the two lists (*M* = 9.1 and *M* = 9.1 for high-familiarity melodies of List A and List B, respectively; *M* = 3.7 and *M* = 4.1 for low-familiarity melodies of List A and List B, respectively).

Recognition Sensitivity Analyses

We once again did two sets of analyses with respect to recognition sensitivity. In our first set of analyses we converted all OT-specific and familiarity-specific hit rates and their OT-nonspecific and familiarity-specific false alarm rates into *d'* scores, disregarding response type. Because list order (high-familiarity list first vs. low-familiarity list first) did not show significant effects, we collapsed all data across this factor. The resulting 2 (OT) × 2 (melody familiar-ity) ANOVA on these *d'* scores revealed only a main effect for melody familiarity, showing that sensitivity levels were higher for high-familiarity tunes (*M* = 1.38, *SD* = 0.64) than for low-familiarity tunes (*M* = 1.11, *SD* = 0.64). All other effects were nonsignificant (*p*s > .10).

For the second set of analyses, all recognition data were converted into *d'* scores, this time using each participant's R, K, and G hit rates for each OT and melody familiarity condition and their respective R, K, and G false alarm rates, thus obtaining R, K, and G *d'* scores across two OT and two melody familiarity conditions (Figure 2).

List order did again not show significant effects (*p*s > .10), so we collapsed all data across this factor. An OT (mood vs. counting) × melody familiarity (high vs. low) × response type (R, K, G) repeated-measures three-way ANOVA on the *d'* rates revealed a main effect for response type, $F(2, 94) = 10.27$, $MSE = 1.15$, $p < .001$, partial $\eta^2 = .18$, indicating that participants' recognition sensitivity was higher for R (*d'* = .79) and K (*d'* = .65) than for G (*d'* = .31) responses.
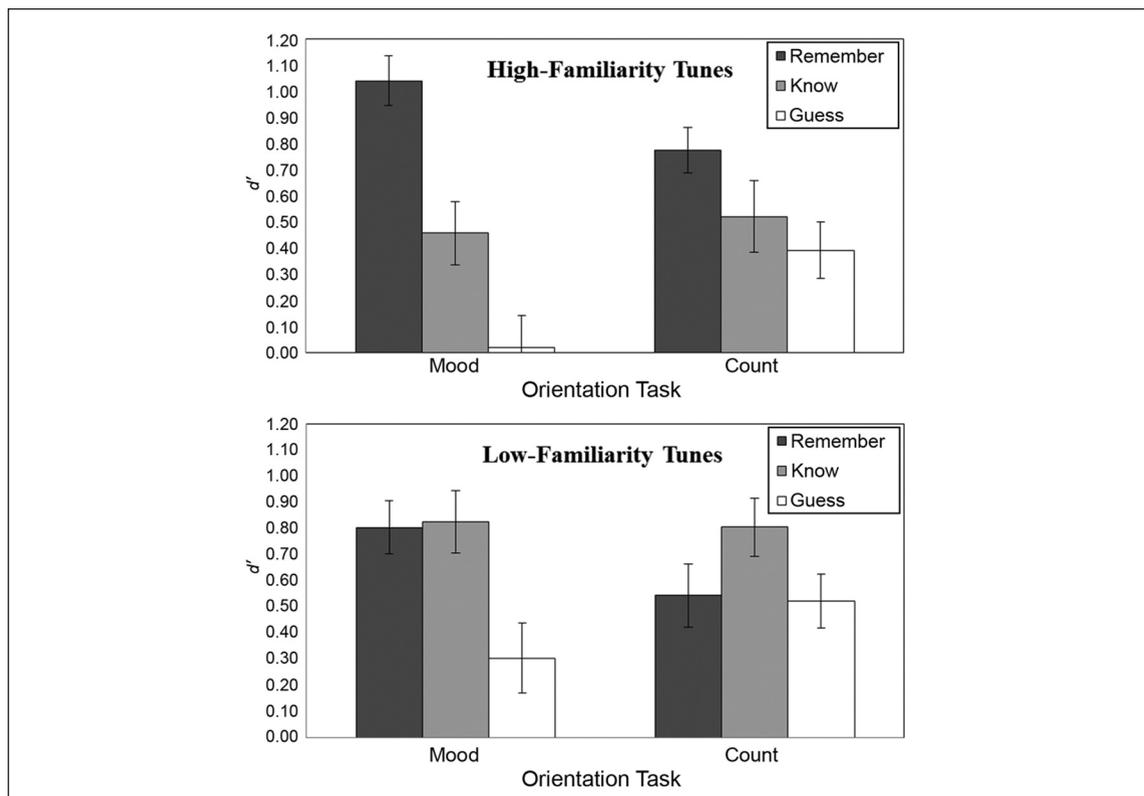

FIGURE 2. *d'* values with *SE*s as a function of orienting task, melody familiarity, and response type, Experiment 2

As in Experiment 1, we also found a significant interaction between OT and response type, $F(2, 94) = 10.27$, $MSE = 0.42$, $p < .001$, partial $\eta^2 = .16$. When we look at Figure 2, we see once again that regardless of melody familiarity both R and G but not K d' scores change greatly across OT conditions. G d' scores again seem to mimic R d' scores but in an almost opposite way. Paired t tests confirmed that the LOP effect in R d' scores was significant for both high- and low-familiarity melodies, $t(47) = 2.73$, $p < .01$ and $t(47) = 2.17$, $p < .05$, respectively. Paired t tests also confirmed a reverse LOP effect in G d' scores for high-familiarity melodies, $t(47) = 2.78$, $p < .01$, and marginally for low-familiarity melodies, $t(47) = 1.84$, $p = .07$.

Furthermore, we found a significant interaction between melody familiarity and response type, $F(2, 94) = 5.56$, $MSE = 0.75$, $p < .001$, partial $\eta^2 = .11$. Paired t tests comparing R and K d' scores showed that for high-familiarity tunes R d' scores were significantly higher than K d' scores ($M = .93$ and $M = .57$, respectively, $t(47) = 2.44$, $p < .05$), whereas for low-familiarity tunes, we observed slightly higher K than R d' scores ($M = .92$ and $M = .78$, respectively), but this difference did not reach significance ($p > .10$). All remaining main and interaction effects were nonsignificant (ps > .10).

Recognition Bias Analyses
A 2 (OT) × 2 (melody familiarity) ANOVA on c scores using OT- and familiarity-specific hit rates and OT-nonspecific, familiarity-specific false alarm rates, disregarding response type, revealed an effect of melody familiarity, that is, high-familiarity melodies led to a "no" bias ($M = 0.28$), whereas low-familiarity melodies led to a "yes" bias ($M = -0.28$), $F(1, 47) = 59.20$, $MSE = 0.26$, $p < .001$, partial $\eta^2 = .56$.

A 2 (OT) × 2 (melody familiarity) × 3 (response type) ANOVA on c scores using OT- and familiarity-specific R, K, and G hit rates and their respective R, K, and G false alarm rates revealed again a familiarity main effect, $F(1, 47) = 51.14$, $MSE = 0.30$, $p < .001$, partial $\eta^2 = .52$, showing that high-familiarity melodies led to a lesser "yes" bias ($M = -1.11$) than low-familiarity melodies ($M = -1.41$). There were also two significant interaction effects: an interaction between OT and response type, $F(2, 94) = 8.94$, $MSE = 0.11$, $p < .001$, partial $\eta^2 = .16$, and one between familiar-

ity and response type, $F(2, 94) = 29.00$, $MSE = 0.51$, $p < .001$, partial $\eta^2 = .38$. With respect to the first interaction, we observed that the mood task resulted in a lower "yes" bias in R responses than the counting task ($M = -1.19$ and $M = -1.32$, respectively); no such change was observed in K responses ($M = -1.37$ and $M = -1.36$, respectively). With respect to the second interaction, we observed a particularly low "yes" bias in R responses for high-familiarity tunes ($M = -0.80$) compared with low-familiarity tunes ($M = -1.71$). "Yes" bias differences between high- and low-familiarity tunes were less dramatic for K ($M = -1.24$ and $M = -1.50$, respectively) and G ($M = -1.30$ and $M = -1.10$, respectively) responses.

DISCUSSION

We found that the mood judgment task compared with the counting long notes task yielded higher recognition sensitivity rates of R responses not only for high-familiarity (1.04 vs. 0.77, respectively) but also for low-familiarity tunes (0.80 vs. 0.54, respectively). We also replicated the curious "reverse" LOP effect in G responses, which we observed in Experiment 1, that is, the counting task yielded higher recognition sensitivity in G responses than in the mood task, particularly for high-familiarity melodies. It is likely that R and G responses act in antagonism because if LOP manipulations work on R- but not K-type recognition sensitivity, then whatever manipulation boosts R sensitivity should be expected to necessarily decrease G sensitivity, which reflects guessing behavior.

We were slightly surprised by the fact that even though the recognition sensitivity rates of R responses for the mood task were almost identical for Experiments 1 and 2 (1.05 in Experiment 1 and 1.04 in Experiment 2), those for the counting task were quite different (0.45 in Experiment 1 and 0.77 in Experiment 2). It could be that the counting task in the context of three other tasks resulted in less distinctive processing than when engaged in the presence of only one other task.

Finally, our findings regarding the effect of melody familiarity on R and K responses were in line with the literature. Gardiner and Radomski (1999) have shown that listeners report more R than K responses for familiar tunes and vice versa for unfamiliar tunes. We, too, found higher R than K d' scores

for high-familiarity tunes and a reverse trend for low-familiarity tunes. Because we found an LOP effect in R responses for both high- and low-familiarity melodies, it appears that conceptual processing did at least partly counteract the lowering of R responses due to low familiarity.

We once again observed an overall "no" bias when looking at general "yes" responses while disregarding R, K, and G judgments and an overall "yes" bias when looking at the "yes" responses together with their R, K, and G judgments. High- compared with low-familiarity tunes produced a "no" bias in general recognition responses and a lower "yes" bias in R-type recognition responses. Both independent variables, OT and melody familiarity, exerted their influences primarily on R rather than K response biases.

GENERAL DISCUSSION

Our two studies found a consistent effect of OT on R-type recognition sensitivity for both high- and low-familiarity tunes. In Experiment 1, we found that the mood task was singularly effective in creating higher R d' rates compared with any of the other tasks. In Experiment 2, we replicated this mood effect, and, more interestingly, we found that it occurred for both high- and low-familiarity tunes.

Our main goal was to take a closer look at LOP effects in episodic melody recognition by using finer measures of recognition. A large body of literature on recognition memory suggests that recognition judgments are based partly on recollective awareness of a previously encountered event, when people actually remember the episode of having encountered that event, and partly on a sense of familiarity, when a particular item among a set of targets and lures simply seems to stand out as an old item (Gardiner, 1988; Tulving, 1985). LOP is cited as one of the prominent variables that affects the first but not the second component. A semantic OT, for example, increases the proportion of recollection-based recognition, whereas a perceptual OT does not. Furthermore, neither task affects familiarity-based recognition judgments (Gardiner, 2001; also see Wagner & Gabrieli, 1998). In his extensive 2002 review, Yonelinas concluded that the two components, recollection and familiarity, appear to be independent and that recollection is linked to a threshold-type retrieval process whereas familiarity

appears to be linked to more signal detection–type processes (cf. Donaldson, 1996). Yet other researchers believe recollection and familiarity to be guided by a single signal detection process (e.g., Dunn, 2008; Rotello, Macmillan, Hicks, & Hautus, 2006). However, this controversy is beyond the scope of this article because our primary interest was to simply see whether LOP manipulations in music show differing effects on R and K sensitivities to begin with.

Studies using LOP and music have so far looked only at overall measures of recognition sensitivity. For the very reason that LOP manipulations seem to affect one but not the other of the two components of recognition, we believe that the use of "remember" and "know" judgments is crucial to the understanding of why LOP effects have been so elusive with musical materials. As mentioned earlier, "remember" experiences in memory are believed to be influenced by manipulations that affect an item's distinctiveness, whereas "know" responses are believed to be influenced by manipulations that affect an item's perceptual fluency (Rajaram & Geraci, 2000). We therefore expected that our LOP manipulations using music would similarly show their effects on participants' R responses rather than their K responses. In Experiment 1, we used four OTs, two conceptual and two perceptual ones. Because conceptual tasks are defined as tasks that require interpretive processes in the sense of going beyond the immediately given, we expected those tasks to increase item distinctiveness, compared with our perceptual tasks that simply asked participants to process certain physical aspects of the tunes. Our findings showed that, compared with the other tasks, the mood task produced the highest recognition sensitivity in R responses, which might suggest that this task was unique in increasing item distinctiveness.

In addition, because so far no LOP effects have been found with tunes that were not familiar, we were curious as to whether this usual null effect might be due to the difficulty of establishing distinctiveness with such materials. Results of Experiment 2 revealed that the mood task led to a higher recognition sensitivity in R responses with low-familiarity tunes just as much as with high-familiarity tunes. Nevertheless, not finding an LOP effect at the level of overall recognition sensitivity with low-familiarity musical stimuli (cf. Peretz, Gaudreau, & Bonnel, 1998; Halpern & Müllensiefen,

2008) no longer comes as a surprise. Because low-familiarity items, be they words (Gardiner & Java, 1990) or music (Gardiner & Radomski, 1999), are known to produce fewer R-type responses, in the case of music we would have to come up with an LOP manipulation that is strong enough to override this disadvantage of low-familiarity tunes compared with high-familiarity tunes. Our two studies also showed that even with high-familiarity tunes, LOP manipulations, though sufficient to show significant effects on recognition sensitivity in R responses, may not be effective enough to show differences at a coarser level of measurement, such as overall recognition sensitivity.

Studies have shown that it is indeed not so much the conceptual–perceptual component but the distinctiveness component that is crucial in whether we see an increase or decrease in recollection-based recognition (Mäntylä, 1997; Rajaram & Geraci, 2000). The distinctiveness component is especially crucial in our case because our lures shared basic features with the targets. Thus, from a distinctiveness perspective, the continuation task, in which participants had to think of the next two or three notes of high-familiarity tunes, may not have produced as distinctive a processing as the mood task. The counting task, on the other hand, merely asked participants to count the long notes in a tune, which is a rote task unlikely to produce elaborate, distinctive processing. However, it is unclear how the contour tracing task should be conceptualized. Are we increasing the tune's distinctiveness by tracing its pitch contour? Or are we increasing its perceptual fluency, or both? Our R and K d' rates do not suggest an increase in perceptual fluency because otherwise, compared with the other three tasks, we should have seen a boost in K d' rates for the tunes processed with the contour task. An interesting future study would be to use a set of OTs that are either all conceptual or all perceptual but differ in their distinctiveness.

1. Target–lure pairs were established with an effort to maintain a certain level of difficulty at test. For that reason, even if they were from different pieces, they shared some musical characteristics, such as both being popular dance pieces played at weddings ("Hava Nagila" and a Turkish folk dance piece called "Kasap Havasi").

2. When we grouped the tasks in terms of the conceptual versus perceptual differentiation for a 2 (OT) × 3 (response type) ANOVA, an OT main effect emerged, $F(1, 31) = 10.26$, $MSE = 0.09$, $p < .01$, partial $\eta^2 = .25$, showing that the conceptual tasks (mood and continuation) led to significantly higher d' levels than the perceptual tasks ($M = 1.57$, $SD = 0.64$ and $M = 1.18$, $SD = 0.58$, respectively). This finding was in line with Peretz et al.'s (1998) LOP finding with familiar tunes. Other than that, there was neither a response type main effect nor an OT × response type interaction.

3. Earlier pilot studies revealed extreme false alarm rates for high-familiarity melodies when presented in a mixed list of high- and low-familiarity melodies, so we decided to manipulate melody familiarity between lists.

REFERENCES
Bower, G. H., & Karlin, M. B. (1974). Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology, 103,* 751–757.

Bowers, J. S. (1994). Does implicit memory extend to legal and illegal nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 534–549.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 521–533.

Craik, F. I. M. (2002). Levels of processing: Past, present . . . and future? *Memory, 10,* 305–318.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Craik, F. I. M., & Watkins, M. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 559–607.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24,* 523–533.

Duchek, J. M., & Neely, J. H. (1989). A dissociative word-frequency × levels-of-processing interaction in episodic recognition and lexical decision task. *Memory & Cognition, 17,* 148–162.

Dunn, J. C. (2008). The dimensionality of the remember–know task: A state-trace analysis. *Psychological Review, 115,* 426–446.

Friedman, A., & Bourne, L. E., Jr. (1976). Encoding the levels of information in pictures and words. *Journal of Experimental Psychology: General, 105,* 169–190.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Gardiner, J. M. (2001). Episodic memory and autonoetic consciousness: A first-person approach. *Philosophical Transactions of the Royal Society B: Biological Sciences, 356,* 1351–1361.

Gardiner, J. M., Gregg, V. H., & Karayianni, I. (2006). Recognition memory and awareness: Occurrence of perceptual effects in remembering or in knowing depends on conscious resources at encoding, but not at retrieval. *Memory & Cognition, 34,* 227–239.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18,* 23–30.

Gardiner, J. M., & Radomski, E. (1999). Awareness of recognition memory for Polish and English folk songs in Polish and English folk. *Memory, 7,* 461–470.

Gregg, V. H., & Gardiner, J. M. (1994). Recognition memory and awareness: A large effect of study–test modalities on "know" responses following a highly perceptual orienting task. *European Journal of Cognitive Psychology, 6,* 131–147.

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

Hyde, T. S. (1973). Differential effects of effort and type of orienting task on recall and organization of highly associated words. *Journal of Experimental Psychology, 79,* 111–113.

Hyde, T. S., & Jenkins, J. J. (1973). Recall for words as a function of semantic, graphic, and syntactic orienting tasks. *Journal of Verbal Learning and Verbal Behavior, 12,* 471–480.

Jacoby, L. L., Craik, F. I. M., & Begg, I. (1979). Effects of decision difficulty on recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 18,* 585–600.

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition, 10,* 33–42.

Mäntylä, T. (1997). Recollections of faces: Remembering differences and knowing similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1203–1216.

Marks, W. (1991). Effects of encoding the perceptual features of picture on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 566–577.

Meyer, L. B. (1956). *Emotion and meaning in music.* Chicago, IL: University of Chicago Press.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48,* 115–138.

Peretz, I., Gaudreau, D., & Bonnel, A.-M. (1998). Exposure effects on music preference and recognition. *Memory & Cognition, 26,* 884–902.

Rajaram, S. (1998). The effects of conceptual salience and perceptual distinctiveness on conscious recollection. *Psychonomic Bulletin & Review, 5,* 71–78.

Rajaram, S., & Geraci, L. (2000). Conceptual fluency selectively influences knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1070–1074.

Repp, B. (1992). Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei." *Journal of the Acoustical Society of America, 92,* 2546–2568.

Rotello, C. M., Macmillan, M. A., Hicks, J. L., & Hautus, M. J. (2006). Interpreting the effects of response bias on remember–know judgments using signal-detection and threshold models. *Memory & Cognition, 34,* 1598–1614.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31,* 137–149.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26,* 1–12.

Wagner, A. D., & Gabrieli, J. D. E. (1998). On the relationship between recognition familiarity and perceptual fluency: Evidence for distinct mnemonic processes. *Acta Psychologica, 98,* 211–230.

Warker, J. A., & Halpern, A. R. (2005). Musical stem completion: Humming that note. *American Journal of Psychology, 118,* 567–585.

Warrington, E. K., & Ackroyd, C. (1975). The effect of orienting tasks on recognition memory. *Memory & Cognition, 3,* 140–142.

Wig, G. S., Miller, M. B., Kingstone, A., & Kelley, W. M. (2004). Separable routes to human memory formation: Dissociating task and material contributions in the prefrontal cortex. *Journal of Cognitive Neuroscience, 16,* 139–148.

Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General, 130,* 361–379.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46,* 441–517.