Key generalization of recognition memory for melodies

Abigail Lincoln Kleinsmith
University at Albany, State University of New York, akleinsmith@albany.edu

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Cognitive Psychology Commons, and the Music Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/1426

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
KEY GENERALIZATION OF RECOGNITION MEMORY FOR MELODIES

by

Abigail L. Kleinsmith

A Thesis
Submitted to the University at Albany, State University of New York
in Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

College of Arts and Sciences
Department of Psychology
2015
Author Note

This research was conducted under the supervision of Dr. W. Trammell Neill, with additional support and guidance from Jeff Bostwick and George Seror III. I am also grateful to my second committee member, Dr. Ronald Friedman, for his invaluable feedback and support. I am indebted to my dedicated research assistants for their support and assistance with data collection over the past three years: Daniela Azaraev, Enzo Belli, Yadi Chen, Sabrina Cox, Caity Gallagher, Roy Harding, Allyson Kaczmarek, Joslin Kennedy, Alec Lewis, Camila Osses, Marisa Serchuk, Mehwish Shahid, Natalie Slavova, Alexis Towler, Nicole Vanderjagt, and Kaelyn Wood.

Collection of the data reported in Experiment 3 was supported by funding provided by the Initiatives for Women Program at the University at Albany, SUNY, in the form of an Initiatives for Women Endowment award.
Abstract

People easily recognize a melody in a previously unheard key, but they also retain some key-specific information. I tested the hypothesis that individuals compare novel melodies to a memory “prototype” representing the central tendency of experienced exemplars. Participants were familiarized with a monotonic eight-note melody in two closely separated keys and tested for discrimination of that melody from others. Test and foil melodies included ones that were the “average” of pitch heights and ones that were more distant in pitch height. Hit rates and discriminability ($d'$) were better for physically closer keys than for harmonically related keys. In follow-up experiments, the same paradigm was investigated with unfamiliar stimuli (whole- and quarter-tones) and with a longer retention interval between initial and final test (48 hour delay). The results suggest that melody recognition depends on proximity to pitch-specific memories rather than an abstract prototype and question the common assumption that key abstraction is crucial to memory for melodies.
Key Generalization of Recognition Memory for Melodies

Generalization from one musical key to another is an important component of memory for melodies. It is necessary that individuals maintain some mental representation of a melody in order to be able to recognize or identify deviations from that melody; a modulation to a new musical key is recognized only if one remembers the previous key. However, to date, there is no definitive answer to the question, “What are the specific components of a melody that facilitate recognition of that melody?”

The ability to recognize a melody in a previously unheard key is a well-documented phenomenon in the music cognition literature (Cuddy & Cohen, 1976; Cuddy, Cohen, & Miller, 1979; Cuddy, Cohen, & Mewhort, 1981; Miyazaki & Rakowski, 2002). Our ability to recognize melodies transposed into novel keys suggests that individuals may store information about musical intervals rather than specific pitches in memory and may, in turn, use information about these intervals to inform recognition. Cuddy and Cohen (1976) presented participants with a standard melody against which they were to judge two comparison melodies, one of which was a correct transposition of the standard and one of which was an incorrect transposition of the standard; a “correct” transposition retains the original melody’s intervals and contour while altering the absolute pitch height of the tones within the melody. Participants’ task was to indicate which comparison melody was the correct transposition of the standard. Cuddy and Cohen (1976) reported that both musically trained and untrained participants were able to accurately recognize brief (three-note) transposed melodies; both groups’ mean accuracies on this task were statistically above chance, despite the musically-untrained group’s performance being “disappointingly poor” (Cuddy & Cohen, 1976, p. 267). In later papers with additional colleagues, these authors reported similar findings both when varying the contour of the standard
melodies (Cuddy, Cohen, & Mewhort, 1981) and when comparison melodies were transposed to either the tritone or the dominant of the tonic pitch of a standard melody (Cuddy, Cohen, & Miller, 1979).

Miyazaki and Rakowski (2002) investigated differences in transposition abilities of possessors and non-possessors of absolute pitch (AP). A standard melody was visually presented to participants and a comparison melody was presented aurally; participants were required to indicate, on a six-point scale, their degree of confidence that the comparison melody was either the same or different from the standard. The standard melody was either tonal or atonal, and the comparison melody was either non-transposed, transposed four semitones downward, or six semitones upward. While AP is not directly germane to the current discussion, it is nonetheless interesting that the non-AP participants’ mean accuracy was statistically equivalent for all three types of comparison melody and were relatively high (approximately 90% for tonal contexts and approximately 75% for atonal contexts). While it is true that the non-AP participants in this study must have possessed some musical knowledge in order to interpret the standard melody and were thus not musically “untrained” as in Cuddy and Cohen (1976), this finding still stands as evidence that information about the intervals of a melody can be utilized to facilitate recognition.

Although there is evidence suggesting that individuals have the ability to store key-general information about melodies (such as the intervals between notes), other data suggest that individuals do retain key-specific information about a melody, such as the particular quality of a pitch in a scale. Schellenberg and Trehub (2003) demonstrated that participants were able to both correctly identify and correctly reject familiar television-show theme songs that were transposed by either one or two semitones; participants’ mean accuracies were significantly above chance.
performance. Additionally, their performance was not significantly correlated with degree of musical training.

Further support for key-specific memory is reported by Smith and Schmuckler (2008); their participants listened to series of dial tones, some of which were altered either up or down in pitch from the standard tone (440 Hz). Their task was to indicate if a given tone was “normal”, “higher than normal”, or “lower than normal”. Participants produced responses that were significant for both of the altered conditions. The authors’ participants were not trained musicians and the dial tone is an inherently non-musical stimulus, thereby lending support to Smith and Schmuckler’s (2008) assertion that some non-musicians appear to possess a form of absolute pitch.

Data supporting the claims that people are able to remember both key-general (musical intervals) and key-specific (absolute pitches) information indicates that neither mechanism can be solely responsible for musical memory. According to the literature to date, it appears that musical representations in memory may be simultaneously abstract and specific. One possible explanation for musical memory that incorporates both of these elements is that melodies may be compared to a prototype representing the average of all encountered instances.

Posner and Keele’s (1968; 1970) prototype theory served as both a foundation and an impetus for the current set of experiments. Posner and Keele (1968) presented participants with random dot patterns that were generated by distorting a “prototypical” dot pattern; small perturbations of any dot in a given direction would yield a pattern that was visually distinct from the prototype. Their participants were trained to categorize these random dot patterns. Participants were presented with a random dot pattern on a slide and were required to press one of four buttons to indicate their choice as to which category the pattern belonged. After each
trial, participants received feedback about the correct category of that pattern. (Initially, accuracy would be low as the participants attempted to categorize the patterns with no prior knowledge of the difference between the categories, but accuracy would be expected to improve over time with feedback). Participants continued with these training trials until they were able to correctly and sequentially categorize two complete lists of dot patterns.

Importantly, participants were not shown the prototype dot pattern from which the random patterns were generated during the categorization process. After successfully categorizing the distortions, participants’ recognition of various dot patterns was tested. This test included the previously categorized dot patterns, the prototype dot pattern, and new distortions of the same prototype; the latter patterns were presented as a control. Participants accurately categorized both the previously studied old exemplars and their associated nonstudied prototype equally well (87% and 85%, respectively) and categorized both better than the new distortions of the same prototype (73%).

In a follow-up study, Posner & Keele (1970) investigated the impact of delaying the final test on memory for the dot patterns. Half of their participants received an immediate final test and the other half returned one week later for the final test. Participants’ ability to categorize the previously studied old exemplars decayed significantly over time (immediate: 80%; one-week delay: 71%), while their ability to categorize a prototype did not decay over time (immediate: 68%; one week: 66%). These results led Posner & Keele (1970) to conclude that information about the prototype is more stable over time than is information about the studied distortions.

The results reported by Posner & Keele (1968; 1970) indicate that the presence of a prototype in memory may be what facilitates recognition. To my knowledge, there has been no investigation into whether prototypes for melodies are formed in memory. While we tend to only
hear instances of recorded melodies in a single key, a stored prototype may promote recognition of that recording, if it is altered in any way from the originally experienced instance of that recording (e.g., if it is played in a different key, if a listener hears a live performance of the song; if a different musician performs a cover version of the song). Regarding whether an auditory analog of Posner & Keele’s visual prototypes exists, it was hypothesized that during a yes/no recognition memory test, participants would compare novel melodies to a prototype representing an average of the pitch heights (i.e., a central tendency) of the to-be-remembered melodies. It was further hypothesized that participants might form either a “physical distance” prototype or a “harmonic” prototype. A “physical distance” prototype would be a key representing the average of the physical distance in frequency between a pair of keys and would be more akin to the dot-pattern prototype of Posner & Keele (1968). For example, the “physical distance” prototype of the keys C (261 Hz) and D (293 Hz) would be C# (277 Hz), which is the key directly in between C and D. However, harmonic relatedness between a pair of keys may also play a role in melody recognition. A “harmonic” prototype would be a key that is harmonically related to a pair of keys, as per the rules of Western music theory; two keys are considered closely related to the degree that their scales share pitches. For example, the keys of C and D each share six of their seven notes with the key of G; therefore, the “harmonic” prototype of the keys C (no sharps) and D (two sharps) would be G (one sharp).

The experiments presented in this report were primarily designed to investigate to what degree key generalization promotes or mediates recognition memory for melodies. This issue was addressed with two broad questions. The first question concerned the possibility of an auditory analog to Posner & Keele’s (1968) prototype theory. Do individuals rely on some type of “average” melody that is representative of encountered exemplars when making judgments
about previously unheard, novel melodies? Experiments 1A, 1B, 2, and 3 in this report address
this question. The second question concerned the impact of key consistency (or inconsistency) on
recognition memory. To what extent is it important or necessary for a melody to strongly
establish a musical key in order for that melody to be subsequently recognized during a test?
Some researchers (e.g., Krumhansl, 1979) have suggested that establishing a musical key is
important to the development of a strong memory trace for a given melody in that key. What if,
as a result of the nature of certain musical stimuli, it is impossible to establish a specific key for
that melody? How would subsequent recognition memory fare for these melodies? Experiments
4 and 5 in this report were designed to investigate these questions.

Participants in Experiments 1A and 1B were familiarized with one target melody in the
keys of C and D. On a subsequent yes/no recognition memory test, they discriminated their
target melody from foil melodies. The participants’ task was to respond “yes” to their target
melody, irrespective of the key in which it was presented. During this test participants responded
to their target melody in the previously studied keys of C and D, to their target melody in the
previously unstudied keys of C# and G, and to foil melodies in all four of these keys. As
discussed above, the keys of C# and G represent the “physical distance” prototype and the
“harmonic” prototype (respectively) of the pair of keys C and D.

In Experiments 1A and 1B, the key of C# was simultaneously a “physical distance”
prototype of the studied keys C and D and closer in absolute physical proximity to the studied
keys of C and D than was the key of G. In order to reverse this relation between physical
distance to the average and proximity to the studied exemplars, participants in Experiment 2
were familiarized with a target melody in two keys that were more physically disparate from
each other: C and F#. As before, participants discriminated their target melody from foil
melodies (presented in the same key as the target). In Experiment 2, participants responded to their target melody in the previously studied keys of C and F#, to their target melody in the previously unstudied keys of C#, D#, and G, and to foil melodies in all five of these keys. In this experiment, the key of D# represented the “physical distance” prototype of the pair of keys C and F#; the keys of C# and G were close in physical proximity to the keys C and F#, respectively.

Drawing from Posner & Keele (1970), Experiment 3 was designed to assess the impact of a delayed final test on recognition memory. In this experiment, a forty-eight hour retention interval was inserted in between initial and final test. If a musical prototype stored in memory truly facilitates recognition, results similar to those of Posner & Keele (1970) should be obtained: participants’ ability to recognize studied exemplars may decay over time, while their ability to recognize a prototype may remain stable. However, participants may instead demonstrate significant forgetting of both the studied exemplars and the prototype at final test; this result would fail to support a musical prototype being stored in memory.

**Experiment 1A**

**Method**

**Participants.** Fifty-three undergraduate University at Albany students were recruited from the subject pool. Participants who completed the experiment were compensated for their time by receiving credit toward fulfilling a research requirement. Information about the participants’ musical experience was collected with the use of a questionnaire (see Appendix A for questions; see Appendix B for a discussion of these data).

**Stimuli.** The musical stimuli were created using Finale Notepad, a free music writing and music notation software. Six monophonic melodies (1-6) (see Figure 1) were created in four keys
(C, C#, D, G), yielding twenty-four unique melodies. Six different lists of stimuli were created and counterbalanced such that each melody was a “target melody” for a given list and the remaining five melodies were “foils” (e.g., target is “Melody 1”, foils are “Melodies 2-6”; target is “Melody 2”, foils are “Melody 1” and “Melodies 3-6”, etc.). A target melody was always studied in the keys of C and D (see Figure 2). The foil melodies differed from the target by no more than three notes; the first three notes and the last two notes in all melodies were identical. Each melody was represented in all six lists in all four keys. The experiment file was created with E-Prime E-Studio software, version 1.2.

Procedure. Participants were seated individually in front of a computer in a quiet room. They read and signed an informed consent form. They were then presented with a demographic questionnaire. Upon completing the questionnaire, each participant was asked to put a provided pair of headphones on their head; they were then instructed to listen to two versions of the same song as they were presented and to learn them as best as they could. Each participant heard their target melody in the keys of C and D five times each for a total of ten presentations. Participants heard only one of the counterbalanced lists described above, such that different participants were assigned different “target” and “foil” melodies. However, it is not the case that each list was heard an equal number of times across participants (Lists 1-5 were each heard by 9 participants and List 6 was heard by 8 participants, yielding 53 participants). Upon completing the study phase, participants engaged in a four-minute, two-digit addition distractor task; the addition problems were randomly selected from a master list.

After the distractor task, participants engaged in a restudy opportunity prior to the final test. In this phase of the experiment, participants learned to discriminate their target melody in the previously studied keys of C and D from foil melodies in the four keys of C, C#, D, and G.
Participants heard their target melody ten times each in the key of C and the key of D for a total of 20 target trials; they also heard each of the twenty foils (i.e., each of the five different nonstudied melodies in each of the four keys) one time each, yielding a total of 20 foil trials. During this restudy opportunity, participants responded to the question “Is this a version of the song that you studied?” It was emphasized that they should respond “yes” to their target melody, irrespective of the note upon which the melody started (i.e., irrespective of musical key). Participants received feedback after each of their responses. Upon completion of this phase, participants engaged in the same randomized four-minute distractor task mentioned previously.

The last phase of the experiment comprised the final test. During the final test, participants were randomly presented with their target melody and all of the foil melodies in all four keys: C, C#, D, and G. The final test differed from the restudy opportunity in that the final test included presentations of a participant’s target melody in the keys that had not previously been studied (C# and G). Participants heard the twenty foil melodies one time each and one target melody five times each in all four keys, yielding a total of forty target plus foil trials. Participants responded to the same question as before: “Is this a version of the song that you studied?” Upon completing the final test, participants were debriefed and thanked for their time.

**Data Analysis.** Hit rates and correct rejection rates were calculated for each subject for each key. A hit rate is defined as correctly identifying a target melody as the one that was previously studied, and a correct rejection is defined as correctly identifying a lure melody as one that was not previously studied. The data were analyzed by calculating signal detection measures; $d'$, $ln (\beta)$, and $c$. A measure of discriminability ($d'$) was calculated by adding the $z$-score for a subject’s correct rejection (CR) rate to the $z$-score for their hit (H) rate. Due to the fact that so many of the participants responded with perfect accuracy (1.0), a conservative correction
was employed to avoid issues when calculating $z$-scores. Any accuracies of 1.0 were adjusted to 0.99 (had there been any accuracies of 0.0, they would have been adjusted to 0.01) (Macmillan & Creelman, 2005).

Measures of response bias ($ln(\beta)$ and $c$) were also calculated. $ln(\beta)$ is a transformation of the likelihood ratio $\beta$, which is relative to the height of the same distributions for a given $z$-score. Both measures indicate the participants’ biases to respond “yes” vs. “no” in the various conditions. $c$ is a measure of the distance of the criterion from the midpoint between the ‘noise’ and ‘signal + noise’ distributions.

These five measures (H, CR, $d'$, $ln(\beta)$, $c$) were analyzed individually across all four keys using a within-groups repeated measures ANOVA.

**Results**

Table 1 reports mean accuracy and signal detection measures for all four keys. Participants responded more accurately to melodies in the studied keys of C and D than those in the nonstudied keys of C# or G. The hit rates differed significantly among keys, $F(3,156) = 7.501$, $p < 0.0005$, MSe = .016, specifically the keys of C and G, $F(1,52) = 17.79$, $p < 0.0005$, MSe = .018, and the keys of D and G, $F(1,52) = 10.84$, $p = 0.002$, MSe = .017. The correct rejection rates did not significantly differ between any pairs of keys, $F(3,156) = .089$, $p = .956$, MSe = .008.

Differences in $d'$ were significant among keys, $F(3,156) = 6.96$, $p < 0.0005$, MSe = .551. $d'$ differed between the keys of C and G, $F(1,52) = 21.24$, $p < 0.0005$, MSe = .529, C# and G, $F(1,52) = 5.95$, $p = 0.018$, MSe = .685, and D and G, $F(1,52) = 5.71$, $p = 0.021$, MSe = .739. Differences in the measure of $ln(\beta)$ were not significant between any pairs of keys, $F(3,156) =$
.740, \( p = .530 \), \( \text{MSe} = 1.538 \), nor were the differences in the measure of \( c \) between any pairs of keys, \( F(3,156) = 1.51, \ p = .213, \ \text{MSe} = .200 \).

**Discussion**

In Experiment 1A, it was expected that participants would be better able to recognize their target melody in the studied keys of C and D than in the unstudied keys of C# and G; this result was obtained (See Table 1). It was also hypothesized that correct rejection rates should not differ among the four keys; this result was also obtained. In looking for evidence to inform my hypothesis about prototype theory, the most important comparison in this experiment is between the hit rates for the keys C# (the “physical distance” prototype) and G (the “harmonic” prototype).

Overall, participants were better able to identify their target melody in the key of C# (mean hit rate: 88%) than in the key of G (mean hit rate: 83%). As such, there were significant differences between the hit rates of two pairs of keys, but differences between the correct rejection rates were not significant. At first pass, this result suggests that participants may be relying on a “physical distance” prototype when making recognition memory judgments. On a broader level, this result also indicates that participants are in fact retaining some key-specific information about their target melody and that they remember which melody was studied in the initial phase of the experiment. The signal detection data are also interesting; differences in the measure of \( d' \) suggest that participants were able to successfully discriminate between different keys and respond accurately, but did not have a bias to respond either “yes” or “no” irrespective of the question (as indicated by the non-significant differences between keys for measures of \( \ln(\beta) \) and \( c \).
In Experiment 1A, participants responded to a target melody in the key of C# with a higher degree of accuracy than to a target melody in the key of G. However, in the previous experiment, proximity to the average is confounded with absolute pitch height; i.e., melodies in the key of G are higher in pitch height relative to those in the key of C#. In order to determine if the obtained patterns of significance were due to this confound, the melody in the key of G was transposed down one octave; thus, Experiment 1B was created using this new melody in the place of the higher G melody from Experiment 1A.

Experiment 1B

Method

Participants. Thirty-three undergraduate University at Albany students who did not participate in Experiment 1A were recruited from the subject pool. Participants who completed the experiment were compensated for their time by receiving credit toward fulfilling a research requirement. Information pertaining to the participants’ musical experience was collected (see Appendix A).

Stimuli. The stimuli used in Experiment 1B were identical to those used in Experiment 1A, with the exception of substituting the lowered melody in the key of G for the initial melody in the key of G.

Procedure. The procedure used in Experiment 1B was identical to that used in Experiment 1A. However, it is not the case that each list was heard an equal number of times across participants (Lists 1-3 were each heard by 6 participants and Lists 4-6 was heard by 5 participants, yielding 33 participants).

Data Analysis. The same data analysis as that used in Experiment 1A was used for Experiment 1B. Additionally, the data from Experiments 1A and 1B were entered into a
between-groups repeated measures ANOVA in order to determine if there was a main effect of “Experiment”; this allows a determination of whether the pitch height of melodies in the key of G significantly impacted participants’ responses.

**Results**

Table 1 reports mean accuracy and signal detection measures for all four keys. The hit rates differed significantly among keys, $F(3,96) = 5.19, p < 0.002$, MSe = .014, specifically the keys of C and C#, $F(1,32) = 6.087, p = .019$, MSe = .0009, the keys of C and G, $F(1,32) = 10.769, p = .002$, MSe = .018, and the keys of D and G, $F(1,32) = 7.0, p = .013$, MSe = .018. The correct rejection rates did not significantly differ between any pairs of keys, $F(3,96) = .660, p = .579$, MSe = .012.

The $d'$ measure differed significantly among keys, $F(3,96) = 2.03, p = .114$, MSe = .708, specifically the keys of C and G, $F(1,32) = 4.635, p = .039$, MSe = .620, and the keys of D and G, $F(1,32) = 5.329, p = .028$, MSe = .669.

The $ln(\beta)$ measures did not differ significantly between any pairs of keys, $F(3,96) = .898, p = .445$, MSe = 1.061. However, $c$ did differ significantly among keys, $F(3,96) = 1.097, p = .354$, MSe = .153, specifically the keys of C and C#, $F(1,32) = 4.574, p = .04$, MSe = .10, suggesting that participants exhibited a response bias when discriminating a pair of melodies in the keys of C and C#. This difference is not of particular interest to the current report.

The between-groups analysis of Experiment 1A and Experiment 1B indicated that there was no main effect of experiment. Experiment 1A yielded the same qualitative results as Experiment 1B, indicating that it was not the case that the confounding of absolute pitch height with proximity to the average drove the patterns of significance reported in Experiment 1A.
There was not a significant difference in hit rates, correct rejection rates or $d'$ across the two experiments.

As before, the most interesting comparison in this experiment is between the hit rates for the keys C# and G. Again, participants were better able to recognize their target melody in the key of C# (mean hit rate: 86%) than in the key of G (mean hit rate: 81%). Of primary importance, when analyzed combining both Experiment 1A and 1B, recognition of the target melody in the key of C# was significantly better than that of the target in the key of G. This held true for both hits, $t(85) = 2.195$, $p = .031$, and $d'$, $t(85) = -2.544$, $p = .013$. Measures of $ln(\beta)$ and $c$ did not significantly differ between these conditions.

**Discussion**

In both Experiments 1A and 1B, participants correctly responded to a target melody in keys that had not been studied previously (C# and G), indicating that participants are storing at least some key-general information in memory (e.g., the contour of the target melody). Participants also responded to melodies in the key of C# more accurately than to those in the key of G, despite the latter being more closely related to the keys of C and D harmonically. These results imply that it is the physical distance between notes that affects a participant’s memory for melodies and not the harmonic distance between keys. The latter finding would have been supported if participants responded more accurately to target melodies in the key of G than in the key of C#. Due to the fact that there was a lack of evidence for the initial hypothesis about participants forming a “harmonic” prototype, further investigation into this relationship was abandoned.

It is worth mentioning that some researchers have reported that harmonic relatedness between two melodies does indeed have an impact on memory for those melodies (Bartlett &
Dowling, 1980). These authors employed a transposition recognition task, where they presented participants with a standard melody followed by a comparison melody that was either presented in the same key, in a closely harmonically related key, or in a distantly harmonically related key. (For the key of C, the key of G is closely harmonically related and the key of B is distantly harmonically related). Bartlett & Dowling (1980) reported that the more closely harmonically related a comparison melody was to its standard melody, the easier it was for participants to correctly respond to that comparison. This conclusion is in direct opposition to the conclusions of Experiments 1A and 1B; however, there are a few critical differences between the present study and Bartlett & Dowling’s (1980) study that may explain the differing results.

Of primary importance, the procedures of Experiments 1A and 1B and those of Bartlett & Dowling (1980) differ substantially. My procedure addresses long-term memory effects, while Bartlett and Dowling’s addresses short-term memory effects. In their procedure, the standard melody could be different from trial to trial, thereby not requiring participants to retain any long term representation of that melody in memory. In contrast, my participants heard one target melody (i.e., one “standard”) to which all other melodies were compared. This latter procedure clearly requires the use of long-term memory faculties. It is possible that my procedure also tapped into some type of a short-term memory component; participants are receiving feedback after every trial and are thus receiving further training with each trial. However, it is important to realize that participants are receiving feedback equally in all of the different keys in which melodies are presented. In this way, to whatever extent that there is a short-term memory component to my procedure, this would not be expected to produce any difference in performance between musical keys.
Another difference that could help explain the differing results is that the first three notes (and the final two notes) of each melody in Experiments 1A and 1B were held constant within a key for both target transpositions and foil melodies; this was not true of the stimuli used by Bartlett & Dowling (1980). In my experiments, to whatever extent a key is being established for the target melodies, it is being equally established for the foil melodies. In Bartlett and Dowling’s (1980) experiments, it is unclear that a key is being equally established for all melodies; the comparison melodies used in their experiments often began on different notes, even when the comparison was in the same key as the standard. It is likely that these differences in procedure and stimuli between my experiments and those of Bartlett & Dowling (1980) account for the differences in results pertaining to harmonic relatedness.

Despite no evidence for a “harmonic” prototype informing recognition, the results of Experiments 1A and 1B are consistent with the idea that participants rely on a “physical distance” prototype when discriminating target and foil melodies at the time of final test. However, an alternative explanation for these same results could be that melody recognition depends on proximity to a studied exemplar (e.g., “nearest neighbor hypothesis”; Reed, 1972) rather than on a type of physical prototype. Therefore, it is possible that participants responded more accurately to a target in the key of C# than in the key of G simply because C# is physically closer to the studied keys of C and D than is the key of G.

Experiment 2 was designed to distinguish between these two possible explanations for the results of the two previous experiments. In Experiment 2, participants were familiarized with one target melody in the keys of C and F#; these two keys are more physically disparate than were the initially familiarized keys in Experiments 1A and 1B. As a result of this manipulation, the key which represented the “physical distance” prototype of the two studied keys (the key of
D#) was farther away from the two studied keys than was the “physical distance” prototype in Experiments 1A and 1B. Furthermore, two keys were employed that were closer in proximity (i.e., “nearest neighbors”) to the two studied keys but did not represent any kind of average. The key of C# is a “nearest neighbor” to the key of C, and the key of G is a “nearest neighbor” to the key of F# (See Figure 3.) These manipulations made it possible to determine if participants were relying on either a “nearest neighbor” strategy or a “physical distance” prototype to inform recognition judgments on a final test.

Experiment 2

Method

Participants. Thirty-six undergraduate University at Albany students who did not participate in either Experiment 1A or 1B were recruited from the subject pool. Participants who completed the experiment were compensated for their time by receiving credit toward fulfilling a research requirement. Information pertaining to the participants’ musical experience was collected (see Appendix A).

Stimuli. The musical stimuli and experiment file were created in the same way as those used in Experiments 1A and 1B. However for Experiment 2, six monophonic melodies (1-6) were created in five keys (C, C#, D#, F#, G), yielding thirty unique melodies. The target melody was studied in the keys of C and F# (see Figure 3).

Procedure. The procedure for Experiment 2 was nearly identical to that of Experiments 1A and 1B. In this experiment, I created an arbitrary name (“The Silver Stream”) for the target melody and told participants that it was the introduction to a folk song. These measures were enacted in order to facilitate easier understanding of the experiment’s instructions.
The addition of a fifth key created lists of stimuli that were longer than those in the previous experiments. During the restudy opportunity, the ratio of target-to-foil presentations was not equivalent (a minor oversight in programming); target melodies were presented twenty-five times at random while foil melodies were presented twenty times at random. Although participants did not necessarily hear each foil melody an equal number of times, all foils were presented equally randomly across all participants.

During the final test, participants heard all of the target melodies in all five keys, including their target melody in keys that had not previously been studied. However, participants did not hear all of the foil melodies; foils were presented fifteen times at random as in Phase II (due to the same programming oversight). Therefore, they randomly heard fifteen foils in any of the five keys and their target melody five times each in all five keys, yielding a total of forty trials.

**Data Analysis.** The same data analysis was used for Experiment 2 that was used for Experiments 1A and 1B.

**Results**

Table 1 reports mean accuracy and signal detection measures for all five keys. The hit rates differed significantly among keys, $F(4,208) = 11.498, p < .0005, \text{MSe} = .024$: C and C# ($F(1,52) = 7.234, p = .01, \text{MSe} = .016$), C and D# ($F(1,52) = 23.775, p < .0005, \text{MSe} = .035$), C# and D# ($F(1,52) = 8.897, p = .004, \text{MSe} = .036$), C# and F# ($F(1,52) = 4.444, p = .04, \text{MSe} = .02$), D# and F# ($F(1,52) = 21.127, p < .0005, \text{MSe} = .028$), and D# and G ($F(1,52) = 20.511, p < .0005, \text{MSe} = .031$). The correct rejection rates did not differ significantly among keys, $F(4,208) = 1.748, p = .141, \text{MSe} = .020$. 

18
The $d'$ measure significantly differed among keys, $F(4,208) = 5.62, p < .0005, \text{MSe} = 1.128$: C and C# ($F(1,52) = 14.771, p < .0005, \text{MSe} = .763$), C and D# ($F(1,52) = 22.21, p < .0005, \text{MSe} = 1.054$), C and F# ($F(1,52) = 5.09, p = .028, \text{MSe} = .858$), C and G ($F(1,52) = 4.24, p = .045, \text{MSe} = 1.214$), D# and F# ($F(1,52) = 6.96, p = .011, \text{MSe} = 1.085$) and D# and G ($F(1,52) = 6.59, p = .013, \text{MSe} = 1.002$).

The $ln(\beta)$ measure significantly differed among keys, $F(4,208) = 8.08, p < .0005, \text{MSe} = 1.788$: C and D# ($F(1,52) = 11.97, p = .001, \text{MSe} = 2.204$), C# and D# ($F(1,52) = 6.84, p = .012, \text{MSe} = 2.365$), C# and F# ($F(1,52) = 4.24, p = .044, \text{MSe} = 1.33$), D# and F# ($F(1,52) = 22.57, p < .0005, \text{MSe} = 1.814$), and D# and G ($F(1,52) = 22.84, p < .0005, \text{MSe} = 1.935$). The $c$ measure also significantly differed among keys, $F(4,208) = 9.75, p < .0005, \text{MSe} = .276$: C and D# ($F(1,52) = 13.79, p < .0005, \text{MSe} = .371$), C# and D# ($F(1,52) = 11.24, p = .001, \text{MSe} = .346$), D# and F# ($F(1,52) = 28.27, p < .0005, \text{MSe} = .281$), and D# and G ($F(1,52) = 26.95, p < .0005, \text{MSe} = .291$).

Of critical importance, recognition of the target melody in the key of D# was significantly worse than that of the target in the keys of C# and G. This held true for both hits, $t(52) = -4.048, p < 0.0005$, and $d', t(52) = -2.148, p = .036$. Measures of $ln(\beta)$ and $c$ also significantly differed among these conditions.

Discussion

The main goal of Experiment 2 was to test between two potential explanations for the results obtained in Experiments 1A and 1B. Were participants relying on a “nearest neighbor” strategy for judging target and foil melodies during a final test, or were they instead relying on a “physical distance” prototype to inform these judgments? If participants were relying on a prototype, a higher hit rate for target melodies in the key of D# would be expected. However, if
participants were relying on pure physical proximity to a studied target melody, a higher hit rate for target melodies in the keys of C# and G would be expected. The latter result was obtained (mean hit rate for C#: 96%; mean hit rate for G: 90%; mean hit rate for D#: 75%). This result supports the conclusion that participants are relying on physical proximity to a studied target melody during recognition and not on either a “physical distance” or a “harmonic” prototype. Despite the results of the previous experiments failing to provide support for prototype theory, one further issue concerning this question remained.

Posner and Keele (1970) found evidence for the presence of a prototype in memory with a one-week retention interval in between a categorization and recognition memory task. Their participants demonstrated significant forgetting of the old exemplars over the one-week delay; however, their ability to categorize the prototype pattern showed no significant decrease over time.

Posner and Keele’s (1970) results indicated that it is possible to simultaneously have both a prototype and specific exemplars in memory. Therefore, it is reasonable to suggest that participants in the present set of experiments also held both a prototype and specific exemplars in memory. However, it may be the case that representations of the studied exemplars are more immediately available in memory than is a prototype, thereby making it easier to rely on proximity to the studied exemplars as the basis for recognition at a short retention interval. It is also possible that at a longer delay between initial and final test, participants would rely more heavily on a prototype stored in memory. It is expected that memory for the studied exemplars will decay over time; if this occurs and the representation of a prototype perseveres in memory, higher hit rates for a “physical distance” prototype key may be obtained if there is a longer
retention interval between initial and final test. These results would lend support to the idea of an auditory analog to Posner and Keele’s (1970) findings.

In order to investigate this idea, Experiment 3 employed a paradigm identical to Experiment 2, with the critical addition of a 48 hour retention interval in between the restudy opportunity and the final test phase.

Experiment 3

Method

Participants. Twenty-four undergraduate University at Albany students were recruited from the subject pool. Participants who completed the experiment were compensated for their time with a cash reward. Participants received $10.00 for their time ($3.00 after the initial test and $7.00 after the final test). Information pertaining to the participants’ musical experience was collected (see Appendix A).

Stimuli. Stimuli in Experiment 3 were identical to those used in Experiment 2.

Procedure. The procedure for Experiment 3 was nearly identical to that of Experiment 2. The only difference was the 48 hour retention interval between the restudy opportunity and the final test.

Data Analysis. The data analysis was the same as that used for Experiment 2.

Results

Table 1 reports mean accuracy and signal detection measures for all five keys. The hit rates differed significantly among keys, $F(4,92) = 4.78, p = .001, MS_e = .018$: C and C# ($F(1,23) = 5.02, p = .035, MS_e = .013$), C and D# ($F(1,23) = 9.98, p = .004, MS_e = .023$), C# and G ($F(1,23) = 4.45, p = .046, MS_e = .015$), D# and F# ($F(1,23) = 8.51, p = .008, MS_e = .018$), and
D# and G ($F(1,23) = 8.32, p = .008, MS_e = .028$). The correct rejection rates did not differ significantly among keys, $F(4,92) = 1.51, p = .204, MS_e = .015$.

The $d'$ measure significantly differed among keys, $F(4,92) = 3.02, p = .022, MS_e = 1.087$: C and C# ($F(1,23) = 5.66, p = .026, MS_e = 1.382$), C and D# ($F(1,23) = 9.75, p = .005, MS_e = .804$), C# and F# ($F(1,23) = 4.14, p = .054, MS_e = .981$), C# and G ($F(1,23) = 4.27, p = .05, MS_e = .913$), and D# and F# ($F(1,23) = 4.70, p = .041, MS_e = .867$). Neither the $\ln(\beta)$ measure nor the $c$ measure differed significantly among keys.

Of critical importance, recognition of the target melody in the key of D# was significantly worse than that of the target in the keys of C# and G. This pattern of significance held true for both hits, $t(23) = -2.331, p = .029$, but not for $d'$, $t(23) = -.978, p = .338$. Measures of $\ln(\beta)$ and $c$ also significantly differed among these conditions.

**Discussion**

The basic findings of Experiment 3 replicated those obtained in Experiment 2; despite the presence of a 48 hour delay between a restudy opportunity and final test, participants were able to recognize their studied target melody and were able to discriminate it from non-studied foil melodies. Overall (and as before), participants responded more accurately to targets in both the keys of C# and G than to targets in the key of D#. However, the measure of $d'$ did not differ significantly between melodies in the keys of C# and G and melodies in the key of D#, implying that discrimination was not significantly different between this specific pair of keys.

The most important finding of Experiment 3 is that participants do not appear to be storing any information in memory about a prototype. It was hypothesized that as time passed between initial and final test, participants may rely more on a prototypical melody to inform their recognition judgments. This pattern of results was not obtained in Experiment 3. Participants’
mean hit rates were best for the keys that were initially studied (C: 93% and F#: 90%), next best for the keys that were close in physical proximity to the initially studied keys (C#: 85% and G: 93%), and worst for the key representing the physical average, or “prototype” of the initially studied keys (D#: 79%). This is the same pattern of results that was obtained when there was no delay between initial and final test in Experiment 2; this finding disconfirms the hypothesis that evidence for the existence of a prototype would emerge at a longer delay.

Interim Discussion

The results of the first set of experiments in this report clearly indicate an effect of key distance on memory for melodies: that is, the more distant (in physical frequency) a given melody is from the key of a previously studied target melody, the more difficult it is for that melody to be correctly identified as a target melody in a different key. This result was obtained regardless of whether the initially studied target melody was presented in two keys close in physical proximity (Exps. 1A and 1B; keys of C and D) or whether the target melody was presented in two more physically disparate keys (Exps. 2 and 3; keys of C and F#). Additionally, this result was demonstrated both with an immediate final test (Exps. 1A, 1B, and 2) and with a 48 hour retention interval between initial and final test (Exp. 3).

The results of these experiments also reliably demonstrate that recognition memory for melodies is neither facilitated nor mediated by some kind of prototype. Thus, there is no evidence supporting the concept of an auditory analog to Posner and Keele’s prototype theory (1968; 1970). While this disconfirms the initial hypothesis posited at the beginning of this report, it is important to note that these findings provide general knowledge about the specific components of melodies that promote memory for melodies.
With the general question about prototypes for melodies addressed, the final two experiments in this report turn to concerns about the importance of key consistency for recognition memory for melodies. Is it necessary for a melody to strongly establish or suggest a specific musical key in order for that melody to be recognized well on a subsequent memory test? In her seminal paper, Krumhansl (1979) suggested that the answer to this question is “yes”. Her participants engaged in a probe tone task; they were initially presented with a single tone (the “standard” tone) and were subsequently presented with a sequence of notes that either strongly suggested a musical key (diatonic sequence) or did not suggest a musical key (non-diatonic sequence). After this sequence, participants were presented with a second single tone (the “comparison” tone) and were asked to indicate whether this tone was either same or different from the standard tone heard earlier. Her results revealed that participants’ same/different judgments were more accurate if the standard tone was diatonic with respect to the intervening sequence of notes than if it was nondiatomic. This result led Krumhansl (1979) to conclude that the nature of the psychological representation of pitch was tonality-specific. Her conclusions comfortably invite the complementary prediction that the establishment of a musical key is important to developing a strong memory trace, despite the fact that this is not explicitly stated in her report.

Experiments 4 and 5 were designed to investigate memory for melodies composed of a set of pitches that make it difficult (and in some cases, impossible) to establish a musical key. Whole tones (Experiment 4) and quarter tones (Experiment 5) are pitches not frequently heard by Western listeners as they are not components of the typical scales and modes used by Western composers. These pitches, when combined into a musical scale, lack an inherent tonal hierarchy that informs the perception of musical structure. If, as Krumhansl (1979) suggests, key
consistency is indeed a crucial factor that contributes to strong memory traces for melodies, recognition performance should differ between the experiments that use diatonic (Exps. 1A, 1B, 2, and 3) and nondiatonic (Exps. 4 and 5) stimuli. However, if it is possible for a strong memory trace to be established even for melodies composed of elements atypical to Western listeners, recognition performance for nondiatonic stimuli may be similar to performance for diatonic stimuli. In order to address these issues, the same paradigm described in the previous experiments in this report was implemented with melodies atypical to Western listeners.

**Experiment 4**

Whole tone scales are composed solely of whole steps (i.e., pitches that are two keys apart on a standard piano). Every note in a whole tone scale is equidistant from the preceding and succeeding note, thereby making every note equipotential within the scale. This property of the whole tone scale creates a “blurred, indistinct effect” (Kaimen, 2010, p. 376), wherein it is nearly impossible to predict which note may occur next in any given whole tone melody. Therefore, it is impossible to establish a musical key for these melodies and thus, reliance on a tonal hierarchy to inform recognition responses would be unhelpful.

The whole-tone melodies used in Experiment 4 were composed of both intervals and pitches that are frequently heard (and are thus highly familiar) to Western listeners. The ambiguity of a whole tone melody is compounded by the fact that Western melodies are rarely composed using solely whole steps. Because it is inappropriate to talk about these stimuli in terms of a “key” for reasons stated above, the melodies will be referred to by the note on which they begin (e.g., “beginning on a C” rather than “in the key of C”).
Method

Participants. Fifty undergraduate University at Albany students were recruited from the subject pool. Participants who completed the experiment were compensated for their time by receiving credit toward fulfilling a research requirement. Information pertaining to the participants’ musical experience was collected (see Appendix A).

Stimuli. Whole tone stimuli were created by altering the six monophonic melodies used in Experiment 1A and applying a whole tone scale, beginning on C, to the basic contour structure of each melody (See Figure 4). In this way, six monophonic melodies (1-6) were created, beginning on four different pitches (C, C#, D, G), yielding twenty-four unique melodies containing solely whole tone intervals. The studied target melody began on the pitches of C and D.

Procedure. The procedure for Experiment 4 combined the basic procedure of Experiments 1A and 1B with the clarification of instructions of Experiment 2.

Data Analysis. The data analysis used for Experiment 1A was used for Experiment 4.

Results

Table 1 reports mean accuracy and signal detection measures for all four “keys”. The hit rates differed significantly among keys, $F(3,147) = 8.482, p < .0005, MS_{e} = .036$: C and G ($F(1,49) = 8.565, p = .005, MS_{e} = .066$), C# and G ($F(1,49) = 10.623, p = .002, MS_{e} = .046$), and D and G ($F(1,49) = 27.649, p < .0005, MS_{e} = .027$). The correct rejection rates did not differ significantly among keys, $F(3,147) = 1.377, p = .252, MS_{e} = .029$.

The $d'$ measure significantly differed among keys, $F(3,147) = 4.06, p = .008, MS_{e} = 1.054$: C and G ($F(1,49) = 6.56, p = .014, MS_{e} = 1.262$), and D and G ($F(1,49) = 10.282, p =$
The $d'$ difference was marginally significant between the keys of C# and G ($F(1,49) = 3.81, p = .057, \text{MS}_e = 1.085$).

The $ln(\beta)$ measure significantly differed among keys, $F(3,147) = 4.255, p = .006, \text{MS}_e = 1.960$: C and C# ($F(1,49) = 7.91, p = .007, \text{MS}_e = 1.379$), C and G ($F(1,49) = 8.252, p = .006, \text{MS}_e = 2.50$), and D and G ($F(1,49) = 6.28, p = .016, \text{MS}_e = 1.772$). The $c$ measure also significantly differed among keys, $F(3,147) = 6.37, p < .0005, \text{MS}_e = 4$: C and G ($F(1,49) = 9.438, p = .0033, \text{MS}_e = .661$), C# and G ($F(1,49) = 5.36, p = .025, \text{MS}_e = .564$), and D and G ($F(1,49) = 17.11, p < .0005, \text{MS}_e = .324$).

Of critical importance, recognition of a target melody starting on C# was significantly better than recognition of a target melody starting on G. This pattern of significance held true for both hits, $t(49) = 3.26, p = .002$, and $d', t(49) = 1.95, p = .05$. These conditions also significantly differed on $c$.

**Discussion**

The most interesting finding of Experiment 4 is that key generalization appears to occur for melodies that lack a specific “key”. The fact that participants were able to discriminate “keyless” target and foil melodies with a high degree of accuracy speaks to this finding. Furthermore, participants responded more accurately to targets beginning on a C# (mean hit rate: 87%) than to targets beginning on a G (mean hit rate: 73%). Despite the fact that whole tone melodies were recognized somewhat more poorly than the diatonic melodies of Exps. 1A – 3, the same general pattern of results was obtained: participants responded more accurately to whole tone targets that were closer in physical proximity to the studied exemplars, apparently extending this finding to stimuli for which no key can be determined.
As mentioned above, Experiment 5 was also designed to assess if it is necessary that a melody strongly establish a musical key in order for that melody to be successfully recognized on a final test; in Experiment 5, this was assessed with quarter-tone-stimuli, instead of the whole-tone stimuli of Experiment 4.

**Experiment 5**

Quarter tone scales are composed of pitches that are not traditionally found in Western music nor typically produced by Western instruments. Because of this, quarter tones (or the pitches that exist in between any two keys on a standard piano) are highly unfamiliar to Western listeners. As with the whole tone stimuli in Experiment 4, it is impossible to establish a musical key for quarter-tone melodies. The quarter-tone melodies used in Experiment 5 were composed of intervals that are frequently heard by Western listeners, but contained highly unfamiliar pitches. Importantly, establishment of a musical key is impossible for these melodies. As was the case for the melodies in Experiment 4, the quarter tone melodies will be referred to by the note upon which they begin (e.g., “beginning on a C” rather than “in the key of C”).

**Method**

**Participants.** Sixty undergraduate University at Albany students were recruited from the subject pool. Participants who completed the experiment were compensated for their time by receiving credit toward fulfilling a research requirement. Information pertaining to the participants’ musical experience was collected (see Appendix A).

**Stimuli.** The quarter tone stimuli were created by altering the six monophonic melodies used in Experiment 1A by quarter tones; adjusting each pair of successive notes by quarter tones in opposite directions created a distance of at least one semitone in between the notes. This adjustment was counterbalanced across participants. In this way, six monophonic melodies (1-6)
were created, beginning on four different pitches (C, C#, D, G), yielding twenty-four unique melodies composed of quarter tone pitches. The studied target melody began on the starting pitches of C and D.

**Procedure.** The procedure for Experiment 5 was identical to that of Experiment 4.

**Data Analysis.** The data analysis used for Experiment 1A was used for Experiment 5.

**Results**

Table 1 reports mean accuracy and signal detection measures for all four “keys”. The hit rates differed significantly among keys, \( F(3,177) = 18.951, p < .0005, \text{MS}_e = .016 \): C and G (\( F(1,59) = 23.643, p < .0005, \text{MS}_e = .026 \)), C# and G (\( F(1,59) = 20.739, p < .0005, \text{MS}_e = .0426 \)), and D and G (\( F(1,59) = 23.007, p < .0005, \text{MS}_e = .029 \)). The correct rejection rates did not differ significantly among keys, \( F(3,177) = .100, p = .960, \text{MS}_e = .011 \).

The \( d' \) measure significantly differed among keys, \( F(3,177) = 13.058, p < .0005, \text{MS}_e = .687 \): C and G (\( F(1,59) = 21.721, p < .0005, \text{MS}_e = .688 \)), C# and G (\( F(1,59) = 22.694, p < .0005, \text{MS}_e = .709 \)), and D and G (\( F(1,59) = 21.718, p < .0005, \text{MS}_e = .966 \)).

The \( \ln(\beta) \) measure significantly differed among keys, \( F(3,177) = 6.279, p < .0005, \text{MS}_e = 1.506 \): C and G (\( F(1,59) = 10.316, p = .002, \text{MS}_e = 1.924 \)), C# and G (\( F(1,59) = 7.284, p = .009, \text{MS}_e = 1.995 \)), and D and G (\( F(1,59) = 10.553, p = .002, \text{MS}_e = 2.015 \)). The \( c \) measure also significantly differed among keys, \( F(3,177) = 9.559, p < .0005, \text{MS}_e = .218 \): C and G (\( F(1,59) = 14.047, p < .0005, \text{MS}_e = .31 \)), C# and G (\( F(1,59) = 10.23, p = .002, \text{MS}_e = .339 \)), and D and G (\( F(1,59) = 15.76, p < .0005, \text{MS}_e = .288 \)).

Of critical importance, recognition of the target melody starting on C# was significantly better than recognition of those starting on G. This pattern of significance held true for both hits,
Discussion. The findings of Experiment 4 were replicated in Experiment 5. Participants’ mean hit rates for target melodies beginning on a C# (95%) were higher than mean hit rates for targets beginning on a G (82%). This finding demonstrates that participants responded more accurately to quarter tone targets that were closer in physical proximity to the studied exemplars, even further extending the main finding of Exps. 1A - 3 to stimuli that lack a tonal hierarchy. Interestingly, the hit rates in Experiment 5 were higher overall than were the hit rates for the diatonic stimuli of Exps. 1A – 3.

Another interesting aspect of the results of Experiment 5 is that mean hit rates for the quarter-tone melodies generally exceeded those for the whole tone melodies in Experiment 4. This result is suggestive of the idea that familiarity with the unique pitches of a novel melody may not be as important to later tests of memory as is familiarity with the combinations of intervals used in that melody.

The results of Experiments 4 and 5 that were obtained with nondiatonic stimuli answer my initial question about the importance of a tonal hierarchy to memory for melodies that lack such a hierarchy. The results suggest that establishment of a musical key is not nearly as important to recognition memory for melodies as is the physical proximity of a given melody to a studied exemplar. These findings cast doubt on Krumhansl’s (1979) suggestion that establishment of a musical key is important to developing a strong memory trace.

These findings also call into question an assertion made by Ben-Haim, Eitan, and Chajut in their recent paper (2014). These authors report the finding that frequency of exposure to musical pitches affects elements of pitch processing, such that frequently heard pitches are
learned more efficiently than are infrequently heard pitches in an identification task. This assertion directly opposes the current set of results: when discriminating melodies composed entirely of unfamiliar and infrequently heard quarter tones, participants’ mean accuracies actually exceeded those of participants discriminating melodies composed of familiar pitches. These data make it apparent that “lifetime exposure” (Ben-Haim et al., 2014, p. 30) to musical pitches alone cannot account for participants’ high performance on recognition memory tasks.

One of the primary differences between my paradigm and that of Ben-Haim et al. (2014) is that participants in the latter’s experiments engaged in a simple pitch detection task, whereas my participants were required to discriminate entire melodies from one another. Melodies are inherently more contextually rich than are single tones in isolation; it could be that the increased complexity of the stimuli in my experiments is driving the differences in the results discussed here. Another difference is that both Ben-Haim et al. (2014) and I define “frequency of occurrence” differently in our experiments. The pitches that composed my stimuli in Experiments 4 and 5 are almost never used in Western composition (sometimes directly due to the fact that most Western instruments cannot physically produce some specific frequencies), whereas the pitches deemed “low” in “frequency of occurrence” by Ben-Haim et al. (2014) are used in Western composition (e.g., D♭), but are not used as frequently as some other pitches (e.g., C). (It is important to note that the “frequency of occurrence” variable was manipulated extra-experimentally in both paradigms; that is, this variable was not manipulated within the experiment.)

**General Discussion**

At the beginning of this report, questions were posited about two general issues concerning recognition memory for melodies. The first set of questions pertained to the idea that
we may store prototypes in memory that facilitate melody recognition; the second set of questions pertained to the idea that the establishment of a musical key is important to having a strong memory trace for a given melody. The results of the set of six experiments presented here did not provide evidence supporting either of these ideas.

The results of Experiment 1A indicated that participants stored information about musical key which facilitated their ability to correctly distinguish between target and foil melodies. Experiment 1B yielded the same qualitative result, in addition to revealing that the results of Experiment 1A were not due to the fact that physical proximity to the average was confounded with absolute pitch height. In both Experiments 1A and 1B, recognition of target melodies in the key of C# was significantly better than recognition of target melodies in the key of G. This result suggests that participants were neither creating nor storing a harmonic prototype, but may rather be relying a physical prototype to facilitate melody recognition. However, this same result could also be explained by the possibility that melody recognition is dependent upon physical proximity to a studied exemplar and not a physical prototype. Experiment 2 tested between the physical-prototype and physical-proximity explanations for the results of Experiments 1A and 1B.

The results of Experiment 2 suggested that musical memory is contingent on physical proximity to a studied melody rather than to a physical prototype. This conclusion is supported by participants responding more accurately to target melodies in the keys of C# and G than to target melodies in the key of D#. This result directly contradicts the author’s initial hypothesis. Experiment 3 used a nearly identical procedure to Experiment 2 and tested the hypothesis that participants would rely more heavily on a prototype in memory, if they received a final test forty-eight hours after an initial restudy opportunity. In all four of the experiments investigating
prototype theory and memory for melodies in this report, no evidence for prototype theory was obtained. Based on the evidence reported here, it is apparent that physical proximity to a studied melody is more important to memory for that melody than is an average of all the experienced instances of that melody.

Experiments 4 and 5 were designed to address the second set of questions outlined at the beginning of this report, concerning the importance of establishing a musical key to memory for a given melody. In this pair of experiments, the impact of atypical, nondiatonic melodies on participants’ ability to remember those melodies was investigated. In general, whole tone melodies were recognized somewhat more poorly than the diatonic melodies used in Experiments 1A – 3, while quarter tone melodies were recognized as well as these diatonic melodies. Additionally, participants responded more accurately to both whole-tone and quarter-tone targets closer in physical proximity to the studied exemplars, replicating the pattern of results obtained with diatonic stimuli with melodies atypical to Western listeners. The evidence obtained from Experiments 4 and 5 indicate that it is not necessary for a melody to establish a strong sense of key or tonality in order for it to be recognized at the time of final test.

In conclusion, the results of the present experiments contribute to the music cognition literature in three major ways. Primarily, they present evidence that memory for melodies does not depend on the storage of a prototype, either at a short or long retention interval. Secondarily, they suggest that it is not necessary for a melody to strongly establish a musical key in order for it to be well recognized on a subsequent memory test. Finally, all six experiments in this report indicate that, irrespective of whether melodies are diatonic or nondiatonic, physical proximity to a studied target melody is an important factor that significantly contributes to recognition of that melody at the time of a test. To my knowledge, no similar data investigating the issue of key
generalization of recognition memory have been reported. The results reported here open the
door for future investigation into the specific components of melodies that cause them to be
retained in memory.
References


Table 1

Accuracy and Signal Detection Measures for Experiments 1A - 5

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Key</th>
<th>Hit Rates</th>
<th>Correct Rejection Rates</th>
<th>$d'$</th>
<th>$\ln(\beta)$</th>
<th>$c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>C</td>
<td>94%</td>
<td>86%</td>
<td>3.42</td>
<td>-0.36</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>88%</td>
<td>86%</td>
<td>3.16</td>
<td>-0.41</td>
<td>-0.11</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>91%</td>
<td>87%</td>
<td>3.17</td>
<td>-0.11</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>83%</td>
<td>87%</td>
<td>2.77</td>
<td>-0.17</td>
<td>0.01</td>
</tr>
<tr>
<td>1B</td>
<td>C</td>
<td>92%</td>
<td>86%</td>
<td>3.02</td>
<td>-0.39</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>86%</td>
<td>82%</td>
<td>2.88</td>
<td>-0.17</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>90%</td>
<td>85%</td>
<td>3.06</td>
<td>-0.34</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>81%</td>
<td>84%</td>
<td>2.60</td>
<td>-0.59</td>
<td>-0.10</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>93%</td>
<td>90%</td>
<td>3.81</td>
<td>-0.22</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>96%</td>
<td>86%</td>
<td>3.15</td>
<td>0.00</td>
<td>-0.02</td>
</tr>
<tr>
<td></td>
<td>D#</td>
<td>75%</td>
<td>86%</td>
<td>2.87</td>
<td>0.78</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>F#</td>
<td>92%</td>
<td>85%</td>
<td>3.40</td>
<td>-0.46</td>
<td>-0.19</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>90%</td>
<td>84%</td>
<td>3.37</td>
<td>-0.51</td>
<td>-0.18</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>93%</td>
<td>87%</td>
<td>3.50</td>
<td>-0.43</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>85%</td>
<td>80%</td>
<td>2.69</td>
<td>-0.57</td>
<td>-0.20</td>
</tr>
<tr>
<td></td>
<td>D#</td>
<td>79%</td>
<td>85%</td>
<td>2.69</td>
<td>0.04</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>F#</td>
<td>90%</td>
<td>87%</td>
<td>3.27</td>
<td>-0.57</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>93%</td>
<td>83%</td>
<td>3.26</td>
<td>-0.74</td>
<td>-0.28</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>88%</td>
<td>72%</td>
<td>2.75</td>
<td>-0.92</td>
<td>-0.37</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>87%</td>
<td>75%</td>
<td>2.61</td>
<td>-0.26</td>
<td>-0.24</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>90%</td>
<td>75%</td>
<td>2.83</td>
<td>-0.68</td>
<td>-0.35</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>73%</td>
<td>79%</td>
<td>2.17</td>
<td>-0.01</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>96%</td>
<td>90%</td>
<td>3.86</td>
<td>-0.45</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>C#</td>
<td>95%</td>
<td>91%</td>
<td>3.88</td>
<td>-0.33</td>
<td>-0.12</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>96%</td>
<td>91%</td>
<td>4.00</td>
<td>-0.48</td>
<td>-0.17</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>82%</td>
<td>91%</td>
<td>3.15</td>
<td>0.36</td>
<td>0.22</td>
</tr>
</tbody>
</table>
Figure 1. Six monophonic eight-note melodies used as stimuli in Experiments 1A and 1B are presented here in the key of C. The first three notes and the last two notes are identical across all six melodies.
Figure 2. Line graph depicting the contour of one melody in different keys that were utilized in Experiments 1A and 1B. C and D are studied, C# is the “physical distance” prototype, and G is the “harmonic” prototype. Serial note position in the melody is plotted on the x-axis. Frequency in hertz (Hz) is plotted as a log transform on the y-axis.
Figure 3. Line graph depicting the contour of one melody in different keys that were utilized in Experiment 2. C and F# are studied, D# is the physical average of both C and F#, and C# and G are close in proximity to C and F#, respectively. Serial note position in the melody is plotted on the x-axis. Frequency in hertz (Hz) is plotted as a log transform on the y-axis.
<table>
<thead>
<tr>
<th>Starting Pitch</th>
<th>Melody 1: Major Scale</th>
<th>Melody 1: Whole Tone Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td><img src="image" alt="C Scale" /></td>
<td><img src="image" alt="Whole Tone Scale" /></td>
</tr>
<tr>
<td>D</td>
<td><img src="image" alt="D Scale" /></td>
<td><img src="image" alt="Whole Tone Scale" /></td>
</tr>
</tbody>
</table>

*Figure 4.* One of the six monophonic melodies (Melody 1) is displayed in two musical keys (C and D) in a major scale in the second column. Whole-tone transpositions of Melody 1 are displayed in the third column; these transpositions were used as stimuli in Experiment 4. This demonstrates how a whole-tone scale was applied to all of the major scale melodies in all four musical keys (C, C#, D, and G).
Appendix A

Demographic Questionnaire Items

1. Do you have normal (or corrected-to-normal) vision?

2. Do you have normal (or corrected-to-normal) hearing?

3. Are you right handed, left handed, or ambidextrous?

4. Do you have any formal musical training?

5. Have you ever played a musical instrument or performed vocal music?

5A. If so, how many years have you played an instrument or performed vocal music?

6. Have you ever taken any classes in music theory?

6A. If so, how many?

7. Have you ever taken any private music lessons?

7A. If so, how many?
Appendix B

Musicianship Data for Experiments 1A – 5

<table>
<thead>
<tr>
<th>Experiment</th>
<th>N</th>
<th>Number of Musicians</th>
<th>Number of Non-musicians</th>
<th>Hit Rate</th>
<th>Correct Rejection Rate</th>
<th>d’</th>
<th>ln(β)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>53</td>
<td>14</td>
<td>39</td>
<td>.978</td>
<td>.716</td>
<td>.914</td>
<td>.634</td>
</tr>
<tr>
<td>1B</td>
<td>33</td>
<td>12</td>
<td>21</td>
<td>.569</td>
<td>.910</td>
<td>.495</td>
<td>.427</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>16</td>
<td>37</td>
<td>.258</td>
<td>.420</td>
<td>.342</td>
<td>.582</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>17</td>
<td>7</td>
<td>.121</td>
<td>.005*</td>
<td>.028*</td>
<td>.772</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>16</td>
<td>34</td>
<td>.471</td>
<td>.877</td>
<td>.937</td>
<td>.416</td>
</tr>
<tr>
<td>5</td>
<td>60</td>
<td>20</td>
<td>40</td>
<td>.152</td>
<td>.778</td>
<td>.849</td>
<td>.833</td>
</tr>
</tbody>
</table>

Note: Asterisks (*) indicate significance at a p < .05 level.

For all six experiments in this report, information about participants’ musicianship was analyzed in order to understand the impact that musical training may have on recognition memory. “Formal musical training” was operationalized as a “yes” response to Question 4 on the demographic questionnaire (see Appendix A). Responses to this categorical variable were centered (with “-1” representing a “non-musician” and “1” representing a “musician”) and entered into a repeated measures ANOVA analysis as a covariate for each of the following variables: hit rates, correct rejection rates, d’, and ln(β). Across the five experiments that included an immediate final test, there was no effect of musicianship on overall final test performance.
However, for the experiment that included a delayed final test (Experiment 3), musicianship significantly interacted with correct rejection rates and discriminability ($d'$).

There are numerous reasons why “musical experience” may have had a null effect on overall recognition memory performance in five of the six experiments. A simple explanation could be that formal musical training is unnecessary to perform well on melody discrimination tasks. Similarly, it could also be the case that the task used in these experiments is not sensitive to the benefits of formal music training. It is also possible that because a large majority of Western listeners live within a culture that so frequently exposes them to music in general, even participants who reported that they received no formal musical training are still expert “listeners”.

For the five experiments that included an immediate final test, the sample of participants always contained more “non-musicians” than “musicians”. Furthermore, sample sizes differed between experiments. While it is unlikely that these factors are driving the null effect of musical experience, they are worth noting as a possible source of variance within the data. Given these results, it is not possible to determine which of these factors or combinations of factors contributed to the null effect of musicianship across the majority of the experiments reported here.

It is probable that the significant interactions of correct rejection rates and discriminability with musicianship in Experiment 3 are a result of the delay between the restudy opportunity and final test. Experiment 3 is the only experiment reported here that manipulated this variable. When analyzing the interaction between musicianship and correct rejection rates, it was the case that musicians always had higher correct rejection rates than non-musicians, collapsed across musical key (88% and 75%, respectively). This result suggests that musicians
are better able to correctly identify a foil melody at a delay than are non-musicians. When analyzing the interaction of musicianship and discriminability, it was the case the patterns of data differed for musicians and non-musicians. On the delayed final test, musicians were best able to discriminate melodies in keys that were studied previously (3.53), next best able to discriminate melodies in keys that were close in physical proximity to the previously studied keys (3.32), and least able to discriminate melodies in the key that represented the “physical distance” prototype of the previously studied keys (2.80). Non-musicians showed a slightly different pattern of results: they were best able to discriminate melodies in keys that were studied previously (3.04), next best able to discriminate melodies in the key that represented the “physical distance” prototype of the previously studied keys (2.42), and least able to discriminate melodies in keys that were close in physical proximity to the previously studied keys (1.69). This pattern of results indicates that non-musicians are more likely to utilize a “physical distance” prototype than are musicians at a longer retention interval. At present, the specific reason for this result is unclear; however, it is evident that memory for melodies decays over time differently for musicians and non-musicians. (Note: Appendix C reports the means for musicians’ and non-musicians’ correct rejection rates and discriminability measures.)

Unlike the five experiments reported with an immediate final test, the sample of Experiment 3 contained more musicians than non-musicians. It is also true that Experiment 3 had a relatively low sample size relative to the other five experiments reported here. However, it seems unlikely that these factors led to the significant interactions reported here.
### Correct Rejection Rates

<table>
<thead>
<tr>
<th>Musical Key</th>
<th>Musicianship Status</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Musician</td>
<td>88%</td>
<td>.1544</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>85%</td>
<td>.2183</td>
</tr>
<tr>
<td>C#</td>
<td>Musician</td>
<td>85%</td>
<td>.1327</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>66%</td>
<td>.2740</td>
</tr>
<tr>
<td>D#</td>
<td>Musician</td>
<td>89%</td>
<td>.1201</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>77%</td>
<td>.2103</td>
</tr>
<tr>
<td>F#</td>
<td>Musician</td>
<td>89%</td>
<td>.1201</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>83%</td>
<td>.1339</td>
</tr>
<tr>
<td>G</td>
<td>Musician</td>
<td>91%</td>
<td>.0964</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>63%</td>
<td>.1765</td>
</tr>
</tbody>
</table>

### Discriminability ($d'$)

<table>
<thead>
<tr>
<th>Musical Key</th>
<th>Musicianship Status</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Musician</td>
<td>3.59</td>
<td>1.3225</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>3.28</td>
<td>2.0184</td>
</tr>
<tr>
<td>C#</td>
<td>Musician</td>
<td>3.17</td>
<td>1.3523</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>1.51</td>
<td>1.7791</td>
</tr>
<tr>
<td>D#</td>
<td>Musician</td>
<td>2.80</td>
<td>1.7125</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>2.42</td>
<td>1.4017</td>
</tr>
<tr>
<td>F#</td>
<td>Musician</td>
<td>3.47</td>
<td>1.3589</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>2.79</td>
<td>1.2791</td>
</tr>
<tr>
<td>G</td>
<td>Musician</td>
<td>3.83</td>
<td>1.0188</td>
</tr>
<tr>
<td></td>
<td>Non-Musician</td>
<td>1.87</td>
<td>1.4111</td>
</tr>
</tbody>
</table>