Affective constraints on acquisition of musical concepts: Children's and adults' development of the major–minor distinction

Clarissa A. Thompson and John E. Opfer

Psychology of Music published online 1 August 2012
DOI: 10.1177/0305735612453365

The online version of this article can be found at: http://pom.sagepub.com/content/early/2012/07/30/0305735612453365
Affective constraints on acquisition of musical concepts: Children’s and adults’ development of the major–minor distinction

Clarissa A. Thompson  
University of Oklahoma, USA

John E. Opfer  
The Ohio State University, USA

Abstract

Across cultures and age groups music has a powerful impact on human affective states. We examined the effect of these affective responses on children’s and adults’ ability to label musical excerpts as major or minor. Adults, 10-year-olds, and 5-year-olds rated affective quality of excerpts that differed by mode, tempo, pitch, and excerpt type and then categorized the excerpts by mode. Trial-by-trial assessment of category judgments indicated performance was accurately predicted by subjects’ association of affective valence with musical properties. Specifically, strength of this association prior to training predicted greater category knowledge, types of errors, and age differences in learning.

Keywords

constraints, development, learning, major–minor distinction, microgenetic method, musical cognition

Category learning is often influenced by ‘enabling constraints’ (Gelman & Williams, 1998). Such constraints direct children’s attention to relevant features of the learning task and thereby (1) speed up learning; (2) lead to predictable errors prior to learning; (3) condition transfer of learning; and (4) influence differential receptivity to the types of information that could lead to learning. Although constrained learning has been observed in several domains hypothesized to affect biological fitness (e.g., species-specific song learning in birds; Marler, 1991), evidence also suggests constraints on learning may be considerably more common (see Siegler & Crowley, 1994, for constraints on learning tic-tac-toe strategies, which presumably have no affect on biological fitness). We argue that enabling constraints can both inform educational practice and illuminate how pre-existing biases, speed of acquisition, and transfer of knowledge are related. This premise guided our study of musical concept learning.

Corresponding author:
Clarissa A. Thompson, University of Oklahoma, 455 W. Lindsey St., 705 Dale Hall Tower, Norman, OK 73019, USA.  
Email: cat3@ou.edu
One constraint that may prove influential on learning musical concepts is the automatic affective response to musical properties. Three main lines of evidence suggest some affective responses are likely available prior to learning musical concepts and thereby have potential to influence acquisition of musical concepts. First, music appears to have universal impact on human affect. Music is found in all cultures known to modern anthropologists, and its apparent lack of practical value strongly suggests this universality derives from the pleasure it induces (Pinker, 1997; Huron, 2003; but for a different view on evolutionary origins, see Justus & Hutsler, 2005; McDermott & Hauser, 2006; Trainor, 2006). Second, music has pronounced effects on neural regions specialized for reward/emotion (Altenmüller, Schurmann, Lim, & Parlitz, 2002; Blood & Zatorre, 2001; Menon & Levitin, 2005; Peretz, Gagnon, & Bouchard, 1998; Schmidt & Trainor, 2001). For example, just as the left hemisphere responds to positive visual images and the right hemisphere responds to threatening visual images, the left hemisphere is activated when music evokes positive affect, whereas the right hemisphere is activated when music evokes negative affect (Altenmüller et al., 2002; Schmidt & Trainor, 2001). Finally, many musical features that produce positive affective responses in adults appear to be processed and preferred by infants (Hannon & Trehub, 2005a, 2005b; Trainor, Tsang, & Cheung, 2002). For example, 2-month-old infants prefer consonant musical intervals to dissonant ones (Trainor et al., 2002).

We hypothesized that these automatic affective responses to music might systematically bias children’s and adults’ acquisition of musical concepts much the way other universal biases (e.g., bias to attend to motion) constrain learning in non-musical domains (e.g., animate/inanimate distinction; Opfer & Gelman, 2010). We examined the influence of affective valence (AV) on children’s and adults’ ability to learn to categorize musical excerpts as belonging to the major or minor mode.

The musical category examined – the major–minor distinction – was chosen for two reasons. First, trained musicians and musical novices alike find it difficult to discriminate between major and minor excerpts (Halpern, 1984; Halpern, Bartlett, & Dowling, 1998); therefore the distinction offers an opportunity for investigating learning trajectories. Second, because major and minor modes are judged to have different affective qualities (Dalla Bella, Peretz, Rosseau, & Gosselin, 2001; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Gregory, Worrall, & Sarge, 1996; Hevner, 1935; Kastner & Crowder, 1990), we expected that these qualities would have high psychological relevance in learning the distinction. Psychological relevance of affective states produced by music is exemplified when educators introduce the major–minor distinction in terms of affective qualities (Department of Education, 2004) and when participants with no musical training spontaneously describe major and minor excerpts in affective terms (Hevner, 1935). Specifically, major pieces of music are described as dynamic, having an upward driving force, determining, defining, joyful, excited, bright, clear, sweet, hopeful, strong, and happy. Minor passages are conversely described as passive, having a downward drawing weight, and expressing gloom, despair, sorrow, grief, mystery, longing, obscurity, restlessness, melancholy, mourning, darkness, depression, and being dull, plaintive, and soothing (Hevner, 1935).

Musical definitions of major and minor modes refer only to formal properties of musical passages. Major musical scales are comprised of half steps between the third and fourth scale tones and seventh and eighth scale tones (with the distance of all other consecutive notes being a whole step). Minor musical scales are comprised of half steps between the second and third notes, and usually between the fifth and sixth notes, specifically for natural minor scales (Cayne, 1991). Thus minor mode melodies have a greater proportion of half steps and augmented second intervals than major mode melodies (Krumhansl, 1990; cf. Huron’s [2006]
Thompson and Opfer

3

discussion of scale degree qualia and statistical regularities of scale degrees that may be learned via Western culture, pp. 143–153). Previous research has shown that expert and non-expert listeners made modality inferences with increasing definiteness as the length of the melodic excerpt progressively increased, and did so even in the absence of harmonic cues (Vos & Verkaart, 1999).

One reason the major–minor distinction is defined in terms of formal qualities rather than by affective ones is that affective qualities that characterize opposite modes also characterize opposite tempi (fast/slow) and pitch (high/low) properties. Excerpts with fast tempi or high pitches produce affective responses similar to ones produced by major excerpts; music with slow tempo or low pitch produces a response similar to one produced by a minor excerpt (Burnham, Kitamura, Vollmer-Conna, 2002; Cooper & Aslin, 1990; Gabrielsson & Juslin, 1996; Hevner, 1937; Jackendoff & Lerdahl, 2006; Juslin, 2000; Trainor, Austin, & Desjardins, 2000). Indeed, effect of tempo on affective quality is generally stronger and more reliable than effect of mode (Gagnon & Peretz, 2003). Thus if people were to rely simply on AV in learning the major–minor distinction, tempo and pitch could provide conflicting as well as facilitative cues because not all major pieces are performed at fast tempi and with high pitches and not all minor pieces are performed at slow tempi and with low pitches. In these cases, learners who rely on AV in labeling mode would err by calling fast/high minor excerpts ‘major’ and slow/low major excerpts ‘minor.’

Although affective qualities do not define the major–minor distinction as well as formal properties, attention to affective qualities may nevertheless enable explicit category learning if the learner: (1) is already predisposed to think of music in terms of affective qualities; (2) already roughly aligns mode with affective valence; and (3) has feedback ruling out non-mode based sources of AV (e.g., being told that a happy-sounding, fast minor excerpt is not categorized as major). Once the learner has perceived positive AV, the learner is hypothesized to track conditional probability that AV is due to mode rather than tempo, pitch, or other non-mode information. Additionally, sensitivity to affective qualities should constrain learners in certain ways. First, it should limit students’ abilities to learn from definitions that do not refer to AV (e.g., definitions that refer solely to formal musical qualities, like whole and half steps), and improve students’ abilities to learn from definitions that refer to familiar, emotional labels such as happy and sad. Second, it should lead to errors on problems that pit tempo-produced or pitch-produced AV against mode-produced AV (e.g., slow major pieces or fast minor ones). Third, it should condition transfer of learning such that transfer is highly likely across problems differing in ways irrelevant to AV (e.g., scales vs. melodies) but not very likely across problems that do differ in AV (e.g., excerpts that differ in tempo).

The hypothesis that learning the major–minor distinction would be constrained by participants’ bias toward affective qualities of major/minor excerpts has a number of implications for various aspects of learning. A useful way to characterize aspects of learning is suggested by previous work on cognitive change (Siegler, 1996), which distinguishes between: (1) variability, differences among individuals in the change process; (2) source, causes that set the change in motion; (3) path, sequence of approaches used prior to/following discovery of the correct approach; (4) rate, amount of time/experience separating initial use of a new approach from consistent use of the approach; and (5) breadth, generalization of learning to other problems/contexts. We explain how these general features of cognitive change might characterize learning the major–minor distinction.

First, prior knowledge impacts learning in many domains and often leads individual children to learn more than less knowledgeable peers. In learning the major–minor distinction,
however, an important source of variability in learning may come from individual differences in how much mode is associated with affective qualities. Across many studies, differences in the strength of this association have been observed within and across age groups (Dalla Bella et al., 2001; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Gregory et al., 1996; Hevner, 1935; Kastner & Crowder, 1990). We hypothesized that the strength of the association between mode and affective qualities would facilitate participants’ ability to accurately label excerpts prior to training. We examined the effect of variability in the mode-AV association by contrasting the mode-AV association among ‘experts,’ participants who correctly categorized 75% of excerpts at pre-test, ‘learners,’ participants who correctly categorized at least 75% of scale excerpts after training, and ‘non-learners,’ participants who correctly categorized less than 75% of scale excerpts after training.

Second, in most studies of learning, the source of improved performance is experience, practice, feedback, direct instruction, and requests to explain observations (Siegler, 2006). Given the extensive experience most people have with music of different modalities and the fact that very few people without formal musical training can accurately label excerpts by mode, we did not expect mere experience would be sufficient to correctly label the mode of excerpts. Rather, we assumed learning the mode distinction required explicit instruction. Assuming subjects’ prior bias to attend to affective qualities of music, we expected active ingredients of such instruction would likely include: (1) explicit information associating mode and AV (e.g., major = happy/cheerful; minor = sad/gloomy); and (2) feedback on the correctness of categorizations.

We compared learning across four groups that differed in training regime. Some groups received feedback, others did not; some were given a definition of mode that referred to AV and others received only a formal musical definition of mode. This latter group was of interest because the formal definition of mode actually has higher cue validity but was hypothesized to have less psychological relevance.

Third, although it is theoretically possible that ideal instruction could set learners on a ‘beeline toward advanced competence’ (Siegler, 2006, p. 469), previous studies of learning led us to expect a more protracted path of change. Specifically, learning often involves two independent events: weakening dominant, incorrect approaches and maintaining use of correct approaches rather than regressing to prior ones or getting seduced by novel approaches. If learning is initially constrained by AV, participants would be expected to rely heavily on it when categorizing by mode. Further, because tempo and pitch have even stronger AV than mode (Dalla Bella et al., 2001; Gagnon & Peretz, 2003; Hevner, 1937), we expected participants’ early attempts to label excerpts as major or minor would involve mapping labels onto fast/slow tempi or high/low pitches, respectively.

To examine the path, we identified approaches participants used to categorize excerpts. The ‘mode approach’ involved correctly labeling major excerpts as major and minor excerpts as minor. The ‘tempo approach’ involved (incorrectly) labeling fast excerpts as major and slow excerpts as minor. The ‘pitch approach’ involved (incorrectly) labeling high excerpts as major and low excerpts as minor. The ‘other approach’ involved a grab-bag of other approaches, including but not limited to labeling major excerpts as minor and minor excerpts as major, labeling all excerpts as major, and labeling all excerpts as minor. These approaches were categorized and tracked on a trial-block to trial-block basis to identify approaches participants used prior to/following first use of the mode approach, with the expectation that learners of the mode approach would first use tempo and pitch approaches due to their high AV.
Fourth, microgenetic studies of learning offer a puzzling picture of the rate at which learning occurs. Children often extend newly acquired strategies rather slowly, with direct instruction (Klahr & Nigam, 2004) leading to faster rates of learning and with learning generalizing very narrowly (Siegler, 1996). However, learning is portrayed as more efficient in some domains. For example, in studies of ‘fast mapping’ of novel words to novel categories (Carey & Bartlett, 1978; Markson & Bloom, 1997), learning is depicted as the immediate uptake of novel information, often in a single trial, with little to no forgetting of the novel words or facts and quite broad application of the new word/fact to all token items to which it applies (see Hahn, 2003, for a different view). This rapid rate of learning is not unique to word learning (Markson & Bloom, 1997). For example, representational change in biology and number also occur in a single trial (Opfer & Siegler, 2004, 2007; Opfer & Thompson, 2008; Thompson & Opfer, 2008).

Learning to label musical excerpts by mode, however, seems to involve much more than a simple label to category mapping. Even if teachers explicitly instruct students to label happy-sounding excerpts as major and gloomy-sounding excerpts as minor, subjects would still need feedback on their categorization performance: many excerpts (fast/high) that are not major sound relatively cheerful to learners, and many excerpts (slow/low) that are not minor sound relatively gloomy to learners. We expected accurate labeling of excerpts to be much slower than learning to apply other category labels, and we expected learning to be accomplished most quickly when learners have both explicit instruction and feedback on answers. We observed trial-to-trial changes in accuracy across conditions in which learners received feedback, explicit definitions, both, and neither.

Fifth, the hypothesis that affective qualities impact excerpt categorization also generated predictions about how broadly participants would generalize their learning. It was assumed that scales and melodies would not differ in their affective qualities, and thus participants should transfer learning from scale to melody excerpts. To examine the breadth of learning, we compared the number of correctly categorized scales to the number of correctly categorized melodies after participants completed training.

In the present study category learning in the domain of musical cognition was examined to determine the impact of enabling constraints on children’s and adults’ ability to label excerpts by mode. Across two experiments, adults, 10-year-olds, and 5-year-olds rated the AV of excerpts (i.e., warm vs. cool) that systematically differed in mode (major vs. minor), tempo (fast vs. slow), pitch (high vs. low), and excerpt type (scale vs. melody). Across a pre-test, training, and post-test phase participants categorized excerpts by mode. Trial-by-trial assessment of learning identified the impact of providing participants with definitions and/or feedback. Some participants received AV definitions (i.e., major = happy/cheerful; minor = sad/gloomy), whereas others received definitions referring to the formal structure of musical excerpts.

It was hypothesized that: (1) an important source of individual differences in knowledge of the major–minor distinction would be the extent to which subjects associated mode with AV; (2) learning the mode distinction would require explicit instruction including information associating mode and AV and feedback on correctness of categorizations; (3) formal definitions that fail to reference AV would not lead to much learning; (4) initial attempts to label mode would involve mapping ‘major’ and ‘minor’ onto musical features that share AV of major and minor excerpts; (5) participants would need feedback on categorization performance to abandon tempo and pitch approaches in favor of the mode approach; and (6) learning that occurred from training on scales would generalize to novel melodies.
Experiment 1

The purpose of experiment 1 was to test hypotheses about adults’ ability to learn the major–minor distinction. Adults listened to musical excerpts that varied in tempo, pitch, excerpt type, and mode and then categorized excerpts by mode. Participants received instruction in a factorial design including definitions that referenced the affective quality of major and minor scales and feedback about the accuracy of participants’ categorizations.

Method

Participants. Sixty-four adults (M = 20.25 years, SD = 2.29 years; 52% females) from an introductory psychology course participated in all phases of the experiment because their pre-test scores indicated they did not already possess the major–minor distinction.

Tasks and materials. Participants completed two tasks: (1) affective valence; and (2) mode categorization. In the affective valence (AV) task, participants rated the AV of 16 scales and melodies as warm/sunny (sun icon), cool/cloudy (cloud icon), or neutral (sun covered by cloud icon) on an 11-point Likert scale (Figure 1). Participants were instructed to consider the entire rating scale when characterizing the AV of individual excerpts. In the mode categorization task, participants explicitly labeled computer-generated (piano timbre) scales and melodies by mode (Figure 2). Half of the excerpts began on notes below middle C and were termed ‘low’ and half began on notes above middle C and were termed ‘high.’ Musical excerpts were composed at two tempi in 4/4 time: fast (110 beats per minute) and slow (55 beats per minute). The structure of melodies was equated with the structure of scales since notes were repeated the same number of times in scales and melodies, the midpoint of excerpts centered on the seventh and eighth scale tones, and scales and melodies were resolved at their completion. Experimental trial blocks (Figure 3) consisted of four musical excerpts (i.e., two major, two minor; two fast, two slow; two high, two low). Order of presentation was counterbalanced within each trial block. Excerpt type (i.e., scale vs. melody) was held constant across a trial block.

Design and procedure. The present experiment (Table 1) was composed of the following phases: AV rating, pre-test, training, and post-test. The overall design of the pre-test, training, and post-test phases was a 2 (mode: major/minor) × 2 (tempo: fast/slow) × 2 (pitch: high/low) × 2 (excerpt type: scale/melody) × 2 (feedback in training phase: yes/no) × 2 (definitions in training
Figure 2. Scale and melody example stimuli: (a) major melody; (b) major scale; (c) minor melody; and (d) minor scale.

Figure 3. Sample trial blocks
phase: yes/no) mixed factorial design where feedback and definitions were between subjects variables. Individually tested participants listened to musical excerpts presented on a computer. The experimenter read all slides aloud, controlled presentation rate, and recorded all responses on a data sheet.

**AV rating phase.** Participants rated the AV of 16 musical excerpts on a Likert scale (Figure 1).

**Pre-test phase.** Participants listened to eight major excerpts (four scales/four melodies) and eight minor excerpts (four scales/four melodies) and labeled these 16 excerpts by mode (see Figure 3). Pre-test excerpts were equivalent to excerpts rated during the AV phase and were again categorized at post-test. When participants correctly categorized more than 75% of pre-test excerpts, they were excluded from the remainder of the experiment since they already knew the major–minor distinction.

**Training phase.** Participants listened to 32 musical scale excerpts and categorized these excerpts by mode. Training differed by experimental condition: (a) participants in the feedback + definitions (F+D) group received both feedback and definitions referencing AV; (b) participants in the no feedback + definitions (NF+D) group received no feedback, but did receive definitions referencing AV; (c) participants in the feedback + no definitions (F+ND) group received feedback but no definitions; and (d) participants in the no feedback + no definitions (NF+ND) group received neither feedback nor definitions.

During training participants categorized the eight scale excerpts originally heard during pre-test as well as 24 new scale excerpts. Participants were randomly assigned to training condition (F+D, NF+D, F+ND, NF+ND) and labeled excerpts as major or minor.\(^1\) Half of participants received affective valence definitions (i.e., ‘Major musical scales are happy and cheerful, and minor musical scales are sad and gloomy’), whereas other participants were given a filler statement to control for time-on-task (i.e., ‘Major musical scales are diatonic in structure, and minor musical scales are harmonic, melodic, or natural’). Additionally, half of the participants received corrective feedback (i.e., ‘Yes, that’s major’ or ‘No, that’s minor’), whereas the other participants did not receive feedback. Participants were trained on scales to facilitate acquisition of the major–minor distinction since formal dictionary definitions mentioned structural components of major/minor scales.

**Post-test phase.** Participants labeled 16 musical excerpts by mode without feedback. Post-test excerpts were equivalent to pre-test excerpts.

---

**Table 1. Design for experiment 1 and experiment 2**

<table>
<thead>
<tr>
<th>Experiment 1</th>
<th>Affective valance task</th>
<th>Pre-test</th>
<th>Training</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4 trial blocks</td>
<td>4 trial blocks</td>
<td>8 trial blocks</td>
<td>4 trial blocks</td>
</tr>
<tr>
<td></td>
<td>(scales and melodies)</td>
<td>(scales and melodies)</td>
<td>(scales)</td>
<td>(scales and melodies)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment 2</th>
<th>Affective valance task</th>
<th>Pre-test</th>
<th>Training</th>
<th>Post-text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2 trial blocks</td>
<td>2 trial blocks</td>
<td>4 trial blocks</td>
<td>2 trial blocks</td>
</tr>
<tr>
<td></td>
<td>(scales and melodies)</td>
<td>(scales and melodies)</td>
<td>(scales)</td>
<td>(scales and melodies)</td>
</tr>
</tbody>
</table>

*Note: The training phase in both experiments was composed of trial blocks consisting of scale excerpts only. The same scales and melodies that were heard/rated on the affective valence task were again heard/labeled at pre-test and also at post-test.*
Results and discussion

We used Siegler’s (1996) aspects of cognitive change (e.g., variability, source, path, rate, and breadth) to organize our analysis of adults’ categorization of the major/minor distinction.

Variability. We tested our hypothesis that individuals who most strongly associate affective qualities with mode would show the greatest potential to accurately label major excerpts as ‘major’ and minor excerpts as ‘minor.’ We measured the strength of the association between mode and AV for each participant (AV span for mode, range 0–100%) by taking the absolute value of the difference in AV ratings for major and minor excerpts and dividing by total span of the Likert scale. Next we examined AV spans for experts, learners, and non-learners.

As predicted, expert adults \((n = 19, M = 42\%)\) had higher AV spans for mode than did non-expert adults \((n = 64, M = 20\%), F(1,82) = 25.0, p < .001\). Further, learners \((n = 14, M = 28\%)\) had higher AV spans for mode than did those adults \((n = 50, M = 18\%)\) who did not learn the distinction, \(F(1,63) = 4.57, p < .05\). These findings supported our hypothesis that the strength of the association between mode and affective qualities facilitated participants’ ability to accurately categorize musical excerpts.

Based on previous work (Dalla Bella et al., 2001, Gagnon & Peretz, 2003; Hevner, 1935; Siegel & Siegel, 1977a, 1977b; Smith, Kemler Nelson, Grohskopf, & Appleton, 1994), we hypothesized that tempo, pitch, and mode would influence ratings of AV even for non-expert adults. A tempo (fast/slow) × pitch (high/low) × mode (major/minor) analysis of variance (ANOVA) was conducted on AV ratings of non-experts. As expected, there was a main effect of tempo, \(F(1,63) = 94.4, p < .0001\), indicating that faster \((M = 1.35)\) excerpts were given more positive AV ratings than slower \((M = −1.39)\) excerpts; of pitch, \(F(1,63) = 21.84, p < .0001\), indicating high-pitched \((M = .37)\) excerpts were given more positive AV ratings than low-pitched \((M = −.41)\) excerpts; and of mode, \(F(1,63) = 78.86, p < .0001\), indicating that major \((M = .95)\) excerpts were given more positive AV ratings than minor \((M = −.99)\) excerpts. There was a tempo × pitch interaction, \(F(1,63) = 7.54, p < .01\). To analyze the interaction, we performed a repeated-measures ANOVA on scores that reflected the difference in AV ratings between high and low excerpts. Regardless of tempo, participants rated high excerpts more positively than low ones, but this difference was more pronounced when excerpts were played at a fast tempo \((mean \ difference = 2.20)\) than at a slow one \((mean \ difference = .94)\). Adults who do not yet know how to label musical mode nevertheless had some implicit understanding of modality since they categorized major excerpts as positive and minor ones as negative. However, AV is not perfectly associated with mode because fast/high minor excerpts sound more positive than either slow/low major or slow/high major excerpts (see Figure 4).

Source. We examined whether adults’ categorizations of mode were affected by training experiences, particularly hearing definitions and receiving corrective feedback. We performed a training group \((F+D, \ NF+D, \ F+ND, \ NF+ND) \times\) test phase (pre-test/post-test) ANOVA on the percentage of correct categorizations. Percentage of correct categorizations increased from pre-test \((M = 57\%)\) to post-test \((M = 65\%), F(1,60) = 20.42, p < .0001\). The training group × test phase interaction was significant, \(F(3,60) = 4.24, p < .01\), indicating some training groups learned more than others. Post-hoc tests indicated learning (as measured by pre-test to post-test change scores) in the F+D group was significantly greater than learning in the F+ND group \((mean \ difference = 13\%)\), NF+D group \((mean \ difference = 13\%)\), and in the NF+ND group \((mean \ difference = 16\%)\) \(ps < .05\). Learning across the F+ND, NF+D, and the NF+ND groups did not differ significantly from one another, but proportion of correct answers was only greater
that expected from chance responding in groups that received either definitions, feedback, or both (NF+D, t(15) = 2.26, p < .05, M = 61%; F+ND, t(15) = 2.66, p < .05, M = 60%; F+D, t(15) = 4.95, p < .0001, M = 71%). This suggests that learning of the major–minor distinction did not occur without direct instruction via A V definitions and/or feedback.

To further assess the differential impact of training condition we performed a two-factor ANOVA with relative change (post-test–pre-test/1-pre-test) as the dependent variable (DV) to control for group differences on pre-test. There was a trend towards a main effect of feedback (feedback, M = .26, no feedback, M = .09), F(1,63) = 3.83, p < .10 and definitions (definitions, M = .27, no definitions, M = .08), F(1,63) = 5.28, p < .05. However, there was no significant feedback × definitions interaction (F+D, M = .39, F+ND, M = .12, NF+D, M = .15, NF+ND, M = .04), F(1,63) = .86, non-significant (ns).

Findings indicated that feedback and AV definitions aided correct labeling of excerpts by mode, and explicit instructions associating mode and AV or feedback concerning the correctness of categorizations could induce learning within an experimental session.

Path. We investigated the path of change to determine whether adults made a ‘beeline toward advanced competence’ or whether musical excerpts’ AV influenced accuracy of labeling throughout learning, leading to combinations of musical features (e.g., fast + minor, slow + major) being difficult to label. We first looked at pre-test, training, and post-test accuracy to determine whether accuracy was affected by tempo, pitch, or mode (Figure 5). A 2 (mode: major/minor) × 2 (tempo: fast/slow) × 2 (pitch: high/low) repeated-measures ANOVA with
mode, tempo, and pitch as within-subjects variables was completed on pre-test accuracy scores. At pre-test there was a significant mode × tempo × pitch interaction, \( F(1,63) = 6.68, p < .05 \) as well as a mode × tempo interaction, \( F(1,63) = 18.76, p < .0001 \). To analyze the mode × tempo interaction, we performed a repeated-measures ANOVA on scores that reflected differences in accuracy between fast and slow excerpts. When adults labeled major excerpts, they were more accurate for fast than slow excerpts (\( M = 55\% \)), and when adults labeled minor excerpts, they were less accurate for fast than slow excerpts (\( M = -51\% \)). This finding indicates the association between mode and AV not only enables learning (see ‘Variability of Change’), but also constrains learning by dictating the types of errors that learners make.

Likewise, we examined the mode × tempo interaction, \( F(1,63) = 8.55, p < .001 \). Differences in tempo had a larger effect on accuracy of labeling major (\( M = -36\% \)) than minor excerpts (\( M = 30\% \)). Pre-test categorization accuracy seemed influenced by perceived affective qualities of excerpts: fast excerpts were labeled ’major,’ whereas slow ones were labeled ’minor.’ To determine whether these errors were corrected by feedback, we performed the same analysis for answers at training. There was a significant mode × tempo interaction, \( F(1,63) = 33.55, p < .0001 \), so we performed a repeated-measures ANOVA on scores that indexed difference in accuracy between fast and slow excerpts. When adults labeled major excerpts, they were more accurate for fast than slow excerpts (\( M = 68\% \)); however, when adults labeled minor excerpts, they were less accurate for fast than slow excerpts (\( M = -61\% \)). At post-test there was a significant mode × tempo interaction, \( F(1,63) = 37.03, p < .0001 \), with post-hoc tests indicating that, when adults labeled major excerpts, they continued to be more accurate for fast than slow excerpts (\( M = 78\% \)) and, when labeling minor excerpts, they continued to be less accurate for fast than slow excerpts (\( M = -64\% \)). Throughout learning, prior associations between mode and AV continued to influence learners’ performance, even though accuracy improved considerably.

To examine how gains were achieved, we analyzed various categorization approaches that the participants used. Learners’ approaches on each trial block (TB) were categorized as following the ‘mode’ (i.e., correct), ‘tempo’ (i.e., major = fast; minor = slow), ‘pitch’ (i.e., major = high; minor = low), or ‘other’ approach. We plotted the frequency of each approach

![Figure 5. Experiment 1: adults’ correct categorizations by tempo, pitch, & mode](image-url)
using a backward-trials graphing procedure (Siegler, 1996). TB0 refers to participants’ first use of the mode approach; thus, by definition, use of this approach on TB0 was 100%. TB−1 refers to performance on the TB immediately before TB0; TB+1 refers to performance immediately after TB0.

The most common error prior to first use of the mode approach was categorization by tempo (Figure 6). On TB−1, 48% of adults categorized by tempo, a level that differed reliably from chance, \( t(45) = 4.68, p < .0001 \). After first using the mode approach, adults continued to use it frequently, with most using it on TB+1. \( t(46) = 3.32, p < .01 \) and TB+2, \( t(42) = 2.92, p < .01 \). Many adults also regressed to categorization by tempo. On TB+2, 33% of adults categorized by tempo, a level that differed reliably from chance, \( t(42) = 2.71, p < .05 \), and also on TB+3, 32% of adults categorized by tempo, \( t(40) = 2.54, p < .05 \) (chance: mode = 31%, tempo = 13%, pitch = 13%, other = 44%; Figure 6). These findings supported our hypothesis that participants’ early attempts to label excerpts by mode would involve mapping labels onto fast/slow tempi, which tend to have the same AV as major/minor modes, respectively.

**Rate.** To examine whether different types of training produced different rates of learning, we performed a training condition (F+D, NF+D, F+ND, NF+ND) × trial block (pre-test, training TB1–8, post-test) ANOVA on the percentage of correct scale categorizations. There was a main effect of condition, \( F(3,60) = 4.15, p < .05 \), trial block, \( F(9,52) = 4.09, p < .001 \), and a condition × trial block interaction, \( F(27,162) = 1.42, p < .10 \), indicating that rates of learning differed across training groups. The highest rates of learning were observed in the F+D condition where the average increase in accuracy was 2.11% per trial. The rate of learning was substantially lower in other conditions. In the F+ND condition, there was a 1.11% increase in accuracy per trial, in the NF+D condition, there was a 0.78% increase, and in NF+ND condition there was a 0.11% increase (Figure 7).
To examine how broadly adults generalized their learning, we analyzed whether participants transferred learning from training scales to post-test scales/melodies. There was no statistically significant difference in percentage of correctly categorized pre-test scales ($M = 53\%$) or melodies ($M = 52\%$), $t(63) = .48, ns$, or post-test scales ($M = 62\%$) or melodies ($M = 60\%$), $t(63) = 1.23, ns$. Although participants were trained only on scales, they generalized to novel post-test melodies likely because the surface-level characteristics of the scales and melodies – notes were repeated the same number of times in scales and melodies, the midpoint of the excerpts centered on the seventh and eighth scale tones, and scales and melodies were resolved at their completion – were equated (see Figure 2). One reason scale-to-melody transfer may have been easy is that both the association between AV and mode was equally great for scales and melodies, with the strength of the association differing for fast and slow excerpts. To test this, we compared participants’ AV span for mode on scales to their AV span for mode on melodies, and we found no statistically significant difference in the AV span for mode for scales ($M = 23\%$) and melodies ($M = 21\%$), $t(126) = .77, ns$.

**Experiment 2**

In experiment 2 we tested hypotheses about adults’, 10-year-olds’, and 5-year-olds’ ability to learn the major–minor distinction. It was hypothesized that many (though not all) children would fail to associate AV of major/minor excerpts with mode, leading to relatively fewer children benefiting from AV definitions previously given to adults (experiment 1), but with the same relation between AV and learning otherwise holding for children and adults – in predicting individual differences in learning and errors children would make. Again, participants listened to musical excerpts that varied in pitch and mode and categorized excerpts by mode.
Method

Participants. Sixty-four adults (M = 19.32 years, SD = 1.41 years; 56% females), 64 10-year-olds (M = 10.25 years, SD = 0.54 years; 55% girls), and 48 5-year-olds (M = 5.7 years, SD = 0.49 years; 58% girls) participated in the experiment. Adults were sampled from the same pool of participants as in experiment 1, and children were recruited from nearby public elementary schools.

Tasks and materials. All tasks were the same as experiment 1 except where noted. The chief difference was that tempo varied between subjects rather than within subjects because experiment 1 adults had great difficulty abstracting over tempo differences to label mode, and we wished to simplify the task so children might learn something. Additionally we wanted to examine whether the small effect of pitch on accuracy and low use of the pitch approach among adults in experiment 1 was masked by the presence of tempo. To examine these issues, experimental trial blocks in experiment 2 were comprised of four excerpts (i.e., two major/two minor; two high/two low), with half of the participants labeling all fast-tempo excerpts and the others labeling all slow-tempo ones.

Design and procedure. The experimental design for experiment 2 is shown in Table 1. The design was a 2 (mode: major/minor) × 2 (tempo: fast/slow) × 2 (pitch: high/low) × 2 (excerpt type: scale/melody) × 2 (feedback: yes/no) × 2 (definition: yes/no) mixed-factorial design with tempo, feedback, and definitions as between-subjects variables.

Results and discussion

Variability. As in experiment 1, adult experts (n = 29, M = 42%) more strongly associated affective qualities of excerpts with mode than did non-experts (n = 64, M = 30%), F(1,92) = 7.61, p < .01, and adults (n = 36, M = 34%) who learned the major–minor distinction showed a tendency to have higher AV spans for mode than did adults (n = 28, M = 24%) who did not learn the distinction, F(1,63) = 3.30, p < .10.

To determine whether the strength of the association between mode and affective qualities also facilitated children’s ability to accurately label major/minor excerpts, we performed the same analysis with 10- and 5-year-olds. Because there were too few 10-year-old experts (n = 9) to have the statistical power to generate meaningful tests of our hypothesis, we confined our analysis to learners and non-learners. Ten-year-old (n = 22, M = 38%) learners had higher AV spans for mode than did 10-year-old (n = 42, M = 23%) non-learners, F(1,63) = 7.70, p < .01. Because so few 5-year-olds were either experts (n = 5) or learners (n = 6), we combined the groups to test whether their AV association was greater than for 5-year-old non-learners (n = 44). This analysis showed that 5-year-old experts/learners (M = 3.34) did have significantly higher AV spans for mode than did non-learners (M = 1.9), F(1,53) = 7.01, p < .01.

Association between AV and mode increased with age as indicated by a one-way ANOVA on AV span for mode scores. There was a main effect of age on AV span for mode, F(2,218) = 7.32, p < .001. Post-hoc tests indicated that 5-year-olds associated AV with mode less than adults, t(144) = 3.88, p < .001 and 10-year-olds, t(124) = 2.24, p < .05. Even young children are able to associate positive valence with major excerpts and negative affective valence with minor excerpts, and the strength of these associations increases with age.
Finally, we compared the effect of mode on AV to the effects of pitch by performing a pitch (high/low) × mode (major/minor) repeated-measures ANOVA on AV ratings made by non-expert adults, 10-year-olds, and 5-year-olds (Figure 8). For adults, major excerpts (major + high, M = 1.85, major + low, M = 1.23) received more positive AV ratings than minor excerpts (minor + high, M = −.83, minor + low, M = −1.45), F(1,62) = 77.25, p < .0001. For 10-year-olds, too, major excerpts (major + high, M = 1.65, major + low, M = 1.39) were given more positive AV ratings than minor excerpts (minor + high, M = −.36, minor + low, M = −.6), F(1,62) = 31.31, p < .0001, and 5-year-olds also gave more positive AV ratings to major (major + high, M = .85, major + low, M = .41) than minor (minor + high, M = .03, minor + low, M = −.63) excerpts, F(1,46) = 7.07, p < .05.

As in experiment 1 and measured by AV span for mode, participants reported that major excerpts sounded ‘sunny,’ whereas minor excerpts sounded ‘cloudy;’ however, this tendency was considerably more consistent across fast and slow excerpts in this experiment (cf. Figures 3 and 7), presumably because fast and slow excerpts were rated by different groups of participants and thus the two tempi failed to contrast with each other. Therefore learning to label mode may also be easier because there are fewer exceptions to the rule ‘Major musical scales are happy and cheerful, and minor musical scales are sad and gloomy.’

We also observed pitch differences in AV ratings. Among adults, high-pitched excerpts (M = .44) were rated more positively than low-pitched ones (M = −.11), F(1,62) = 4.16, p < .05. However, among 10-year-olds F(1,62) = .47, ns, and 5-year-olds, F(1,46) = 1.87, p = .18, there was no main effect of pitch, although direction of the effect was similar to that of adults even for the youngest children.

**Source.** To determine whether training had an effect across age groups, we first performed a training group (F+D, NF+D, F+ND, NF+ND) × test phase (pre-test/post-test) ANOVA on percent of correct categorizations separately for each age group. The percentage of correct categorizations for adults increased from pre-test (M = 56%) to post-test (M = 70%), F(1,60) = 23.32, p < .0001. Additionally, training group interacted with test phase, F(3,60) = 11.79, p < .0001. Post-hoc tests indicated that learning in the F+D group (as measured by pre-test to post-test change scores) was significantly greater than learning in the NF+ND group (mean difference = .47, p < .0001) and learning in the F+ND group (mean difference = .2, p < .05). Additionally, learning was significantly greater for the F+ND group than for the NF+ND group (mean difference = .26, p < .01), and learning was significantly greater for the NF+D group than for the NF+ND group (mean difference = .36, p < .0001). Three out of four training groups of adults made mode categorization judgments at above chance levels (F+ND, t(15) = 4.07, p < .01, M = 72%; F+D, t(15) = 8.01, p < .0001, M = 87%; NF+D, t(15) = 4.01, p < .01, M = 74%).

For 10-year-olds, percentage of correct categorizations did not significantly increase from pre-test (M = 54%) to post-test (M = 57%), F(1,60) = .89, ns; however, there was a trend toward a training group × test phase interaction, F(3,60) = 2.40, p < .10. Post-hoc tests indicated learning in the F+D group was significantly greater than learning in the NF+ND group (mean difference = .26, p < .05). Only one training group of 10-year-olds made mode categorization judgments at above chance levels (F+D, t(15) = 2.59, p < .05, M = 69%). For 5-year-olds, percentage of correct categorizations failed to increase from pre-test (M = 48%) to post-test (M = 51%); therefore there were no main effects for test phase or significant training group × test phase interactions for 5-year-olds.

To further assess the differential impact of receiving feedback, definitions, or both during training we performed a two-factor ANOVA with relative change (post-test–pre-test/1-pre-test)
Figure 8. Experiment 2: affective valence ratings of 5-year-olds, 10-year-olds, and adults
as DV. Among adults, receiving feedback produced greater learning than not receiving feedback (feedback, $M = .51$, no feedback, $M = .02$), $F(1,63) = 12.76, p < .001$, and receiving definitions produced greater learning than not receiving them (definitions, $M = .56$, no definitions, $M = -.03$), $F(1,63) = 9.54, p < .001$. Overall, feedback and definitions seemed to have additive effects on learning (feedback, $F(1,63) = 12.76, p < .001$, and receiving definitions, $F(1,63) = 9.54, p < .001$), $F(1,63) = 3.96, p < .10$, although there was a trend towards an interaction. Among 10-year-olds, too, receiving definitions produced more learning than not receiving them (definitions, $M = .56$, no definitions, $M = -.03$), $F(1,63) = 9.54, p < .001$, and feedback tended to produce more learning than not receiving feedback (feedback, $M = .12$, no feedback, $M = -.21$), $F(1,63) = 3.6, p < .10$; providing feedback and definitions did not produce interactive effects, $F(1,63) = .63, ns$.

Among 5-year-olds there was considerably less learning than in the other age groups and considerably less of an impact of training. We failed to find an effect of feedback $F(1,31) = .26, ns$, definitions $F(1,31) = .66, ns$, or interactive effects $F(1,31) = .07, ns$. Given 5-year-olds’ ability to learn from feedback and definitions in other concept learning studies (see Siegler, 2006, for review), the most likely reason for preschoolers’ difficulties in learning to label mode was their relative lack of sensitivity to affective differences between major and minor excerpts. Whether 5-year-olds have cognitive prerequisites for learning musical concepts (e.g., attentional control) or whether our task was so difficult that we observed floor effects, however, is an open question.

These findings indicated that feedback and AV definitions are an important source of adults’ ability to correctly label major and minor excerpts, and findings from adults supported our hypothesis that explicit instructions associating mode and AV or feedback concerning correctness of categorizations could induce learning. Results from children are consistent with this general idea: (1) 10-year-olds also benefited greatly from receiving both feedback and definitions referencing AV; and (2) these conditions were not as effective for 5-year-olds, who did not already associate major/minor scales strongly with positive/negative AV, respectively.

**Path.** We next tested the hypothesis that AV of excerpts would influence accuracy of labeling throughout learning, leading to particular combinations of musical features being difficult to label (i.e., high + minor, low + major). We hypothesized that tempo should cease to interfere with accurate labeling, since it was tested between subjects, and different approaches (e.g., pitch) would dominate behavior prior to/following discovery of the major–minor distinction. To this end, we examined various categorization approaches that participants used throughout learning.

The most common error prior to participants’ first use of the mode approach was categorization by pitch (Figure 9). On TB−1, 57% of adults, $t(31) = 4.84, p < .0001$, 47% of 10-year-olds, $t(29) = 3.63, p < .01$, and 48% of 5-year-olds, $t(20) = 3.10, p < .01$ categorized by pitch, all at above chance levels (chance: mode = 31%, pitch = 13%, other = 56%). After first using the mode approach, adults continued to use it frequently, with most using the approach on TB+1, $t(53) = 7.16, p < .0001$, TB+2, $t(49) = 7.38, p < .0001$, and TB+3 $t(31) = 5.66, p < .0001$. After first using the mode approach, 10-year-olds continued using it, with most using the approach on TB+1, $t(51) = 4.17, p < .0001$, TB+2, $t(44) = 4.64, p < .0001$, and TB+3 $t(33) = 4.99, p < .0001$. Although there were few age differences on TB−1, large age differences were evident on TB+1, where adults were most likely to continue their use of the mode approach ($M = 74%$), 5-year-olds least likely ($M = 35%$), and 10-year-olds in between ($M = 60%$). Age differences were greatest, not in initial approaches, but in continued use of the correct, mode approach. As in experiment 1, these findings supported our hypothesis that participants were expected to make particular errors prior to categorizing by mode (i.e., major = high, minor = low).
Rate. Similar to experiment 1, to examine whether different types of training produced different rates of learning, a training condition (F+D, NF+D, F+ND, or NF+ND) × trial block (pre-test, training TB1−4, post-test) ANOVA on percent of correct scale categorizations was performed for each age separately. For adults there was a main effect of condition, $F(3,60) = 8.38, p < .0001$, trial block, $F(5,56) = 8.2, p < .0001$, and the condition × trial block

Figure 9. Experiment 2: participants’ errors prior to/following ‘mode’ approach: (a) adults; (b) 10-year-olds; and (c) 5-year-olds
interaction was significant, $F(15, 174) = 3.65, p < .0001$, indicating rates of learning differed with experimental condition. Highest rates of learning were observed for the F+D condition where the average increase in accuracy was 6.69% per trial; the rate of learning was substantially lower in other conditions. In the F+ND condition there was a 2.63% increase in accuracy, in the NF+D condition, there was a 4.55% increase, and in the NF+ND condition there was a 2.64% decrease in accuracy per trial. These rates of learning were higher than in experiment 1 since learners were not required to abstract over tempo differences.

For 10-year-olds also there was a main effect of condition, $F(3, 60) = 4.18, p < .01$, trial block, $F(5, 56) = 3.35, p < .01$, and a significant condition $\times$ trial block interaction, $F(15, 174) = 1.76, p < .05$, again indicating that rates of learning differed with experimental condition. The highest rates of learning were observed in the F+D condition where the average increase in accuracy was 2.65% per trial. Rate of learning was substantially lower in other conditions. In the F+ND condition there was a 1.26% increase in accuracy, in the NF+D condition, there was a 1.24% increase in accuracy, and in the NF+ND condition there was a 2.48% decrease in accuracy per trial.

For 5-year-olds, there was a main effect of condition, $F(3, 44) = 3.54, p < .05$, indicating accuracy in the F+D group was greater than in the NF+D group ($mean\ difference = 11.05, p < .05$) and in the NF+ND group ($mean\ difference = 13.91, p < .01$). Five-year-olds in the F+ND group also provided more accurate judgments than children in the NF+D group ($mean\ difference = 10.44, p < .05$). There was no main effect for trial block $F(5, 40) = 1.27, ns$, nor a condition $\times$ trial block interaction, $F(15, 126) = 1.34, ns$. These findings with adults and 10-year-olds supported our hypothesis that accurate labeling of excerpts was quickest by participants who received feedback and who learned about the AV of major/minor excerpts.

**Breadth.** We investigated whether participants transferred learning from training scales to posttest scales and melodies. Again participants generalized learning to novel post-test melodies. Participants in all age groups were just as successful at categorizing scales as they were at categorizing melodies across experimental phases, possibly due to the AV spans for mode not differing based on excerpt type. There was no statistically significant difference in percentage of correctly categorized scales or melodies at pre-test (adults, 59% scales, 59% melodies, $t(63) = .052, ns$; 10-year-olds, 53% scales, 55% melodies, $t(63) = .29, ns$; 5-year-olds, 47% scales, 48% melodies, $t(47) = .84, ns$) or at post-test (adults, 70% scales, 69% melodies, $t(63) = .11, ns$; 10-year-olds, 55% scales, 60% melodies $t(63) = 1.09, ns$; 5-year-olds, 46% scales, 56% melodies, $t(47) = .07, ns$). There was no statistically significant difference in AV span for mode for scales and melodies (adults, 28% scales, 31% melodies; 10-year-olds, 33% scales, 32% melodies; 5-year-olds, 33% scales, 31% melodies). Similar to experiment 1, these findings supported our hypothesis that scales and melodies would not differ in AV and, therefore, even though participants were only trained on scales, learning should transfer to melodies.

**General discussion**

Learning of the major–minor distinction was examined to determine whether category learning in the domain of music – like learning in core domains – might be constrained by domain-relevant properties, specifically by affective qualities that characterize major versus minor musical passages. Across two experiments, dimensions of cognitive change – variability, source, path, rate, and breadth of learning – were examined. We review findings regarding each of these dimensions of change by providing evidence of enabling constraints when children and adults attempted to learn the major–minor distinction, propose a developmental
sequence by which the major–minor distinction might be learned, and consider implications of learning constraints for music education.

First, our examination of variability showed that, among adults, participants who could accurately label the mode of musical passages prior to training were also the participants who were most likely to rate major/minor passages as differing greatly in AV. Interestingly, very few participants could accurately label excerpts prior to training; however, association of mode with AV apparently had a large effect on participants’ ability to learn this distinction. Adult and 10-year-old learners more strongly associated AV with mode than non-learners. Furthermore, the few 5-year-olds who already knew or learned to label excerpts provided higher AV spans for mode ($M = 3.34$) than did non-expert/non-learner 5-year-olds ($M = 1.9$), $F(1,53) = 7.01, p < .01$, indicating that even among 5-year-olds the ability to label excerpts by mode was correlated with the ability to associate excerpts with AV. It is important to point out that learning via analogy (e.g., major = warm, sunny, and happy; minor = cool, cloudy, and gloomy) is not a trivial matter. Analogical reasoning (Gentner, 1983) requires children to draw comparisons across surface-level features and disregard surface-level dissimilarities in an effort to uncover deeper structural similarities. Thus, across all age groups, participants’ association between AV and mode was associated with knowledge or the ability to learn the major/minor distinction.

Second, it was predicted that two sources of information would lead to learning: (1) information associating mode and AV; and (2) feedback on accuracy of categorization performance. Also, descriptions of major/minor mode that did not refer to AV would not lead to fruitful learning.

Adult and 10-year-old participants who received both feedback and definitions during training generally showed the most pre-test to post-test improvement, with the majority of adults providing correct responses after minimal training. Feedback and definitions contributed to correct categorization in an additive manner, indicating neither factor provided necessary and sufficient catalysts for learning.

Training of either type, however, was much less effective in producing learning among 5- and 10-year-olds. Why? Given the correlation between learning and participants’ AV ratings, one reason may be that younger children were less sensitive to major/minor differences in AV and thus could not take advantage of a potent source of learning (e.g., definitions referencing AV). Another reason children failed to learn, however, may be more domain general: failure to repeat correct answers after receiving corrective feedback. Support for this factor came from our path analysis.

Third, we predicted that participants’ initial attempts to label mode would involve mapping the terms ‘major’ and ‘minor’ onto musical features that shared AV of major and minor excerpts. Thus if participants rated fast and/or high excerpts as sunny-sounding and slow and/or low excerpts as cloudy-sounding, their first step on the path toward competence would be to initially label fast and/or high excerpts as major and slow and/or low excerpts as minor.

The path of children’s and adults’ performance was consistent with these hypotheses. During pre-test and training in experiment 1 (but prior to first use of the mode approach), adults labeled fast excerpts as major and slow excerpts as minor, with most individual adults using this ‘tempo approach’ prior to discovery of the correct ‘mode approach.’ In experiment 2, where we reduced salience of tempo (by varying it across rather than between subjects), adults reported large differences in AV for high versus low excerpts (and small differences between fast and slow excerpts), and similarly used the pitch approach most often prior to discovery of the mode approach. Interestingly, the pitch approach was very common among 5- and 10-year-olds, both before first use of the mode approach and after first use of the mode approach (Figure 9).
The most likely reason that training had a smaller impact on 5- and 10-year-olds than adults is that children were more likely than adults to regress to the pitch approach, perhaps due to its greater salience.

Fourth, the general prediction about rate was that participants would need feedback on their categorization performance to abandon tempo and pitch approaches in favor of the mode approach. If so, accurate labeling of excerpts would be accomplished most quickly by participants who were told major excerpts have a positive affective quality, minor excerpts have a negative affective quality, and who were given feedback to abandon tempo and pitch approaches (as in F+D conditions).

Consistently, the rate of learning among adults in experiments 1 and 2 was greatest in the F+D conditions. Ten-year-olds seemed to make use of both feedback and definitions, with the greatest trial-to-trial gains being obtained in the F+D condition. Among 5-year-olds, accuracy was also greatest in the F+D condition; however, regressions common in this group led to little trial-to-trial performance gains and thus a fairly consistent rate of change across conditions.

Fifth, we anticipated that learning which occurred from training on scales would generalize to novel melodies because the scales and melodies were experimentally equated such that differences in AV would be quite low between these excerpt types. According to Barnett and Ceci (2002), the breadth of learning we observed is considered ‘near,’ as opposed to ‘far,’ transfer. Across both experiments, training conditions, test phases, and age groups, participants were equally successful in categorizing scales and melodies. It is an open question as to whether generalization of learning from scales to melodies would have been as robust if more complex melodies containing fewer surface-level commonalities with scales had been used. Future research should aim to investigate whether ‘far’ transfer is promoted after participants receive corrective feedback and definitions that reference the affective valence of the major and minor mode.

Results from the combination of these experiments suggest a four-step developmental framework for acquisition of the major–minor distinction. First, participants must be able to perceive/experience a difference between major and minor musical excerpts (Dalla Bella et al., 2001; Gagnon & Peretz, 2003; Gerardi & Gerken, 1995; Gregory et al., 1996; Hevner, 1935, 1937; Kastner & Crowder, 1990). People might respond to major pieces of music in one way (happiness/cheerfulness) and minor pieces in an affectively opposite way (sadness/dejection).

The present series of experiments established that, during AV rating, participants were able to perceive differences between major and minor excerpts. Average ratings of major excerpts were positive, whereas average ratings of minor excerpts were negative. Although it was a rare occurrence, some participants had an AV span for mode of 0%, indicating that these participants did not perceive AV differences between major and minor excerpts. Non-expert adults had AV spans for mode of 20% and 30% in experiments 1 and 2, respectively; 10-year-olds had an AV span for mode of 28%, whereas 5-year-olds had an AV span for mode of 19% indicating that, on average, participants did perceive differences in AV.

Second, participants must know major excerpts are associated with happy/cheerful AV, whereas minor pieces of music are associated with sad/gloomy AV. In the present experiments, major excerpts were rated positively, and minor excerpts were rated negatively on our AV rating task.

Third, acquisition of the major–minor distinction requires participants to differentiate mode-produced (major/minor), tempo-produced (fast/slow), pitch-produced (high/low), and excerpt type (scale/melody) affective qualities. Path analyses indicated that the most common categorization errors prior to discovery of the mode approach were categorizations based on
tempo and pitch. Participants did not often categorize by excerpt type and percentage of correct categorizations was not significantly different for scales versus melodies. Tempo and pitch variables may have been more salient than mode per se because, even when participants were provided with definitions that referenced affective quality of major or minor excerpts, tempo and pitch errors persisted if participants did not also receive corrective feedback.

Novice listeners in Hevner’s (1935) seminal research described major pieces of music as dynamic, joyful, excited, bright, hopeful, strong, and happy, and minor passages as passive, gloomy, dark, depressed, and soothing. Based on these descriptions it is possible that affective valence is correlated with how much the novice listeners like the excerpts as well as the arousal potential of the excerpts (cf. North & Hargreaves, 1997 adaptation of the circumplex theory of emotion described by Russell, Ward, & Pratt, 1981 and Larsen & Diener, 1992). For example, an exciting and festive excerpt played at a fast tempo would likely be rated high on the dimensions of arousal and likeability regardless of mode. It may be difficult for listeners to tease out arousal, likeability, and emotion in the excerpts used in our experiments.

Fourth, participants must encounter feedback to learn to associate the label ‘major,’ with major excerpts and the label ‘minor,’ with minor excerpts. Despite that, on average, participants observed/perceived differences between major/minor excerpts in terms of their AV ratings, many participants failed to associate correct labels like ‘major’ and ‘minor’ with excerpts they implicitly perceived to differ in AV.

The current results are aligned with previous findings in the domain of categorical pitch perception and implicit learning via statistical regularities. Just as novice listeners found it extremely difficult to categorize pitch or intervals in an absolute manner (Siegel & Siegel, 1977a, 1977b), they also found it difficult to label the mode of excerpts in our present experiments. This difficulty with making correct category judgments likely stemmed from the fact that novices did not initially associate labels like ‘major second’ or ‘minor second’ with the appropriate intervals. Similar to our findings, however, researchers were able to capitalize on familiarity to help novices learn the association between category labels and the intervals to which those labels referred. For example, Smith and colleagues (1994) prompted music novices to associate familiar folk tunes with musical intervals (e.g., ‘Here Comes the Bride’ begins with a perfect fourth, whereas ‘Greensleeves’ begins with a minor third) and these familiar associations allowed novices to more readily perceive the differences between various musical intervals. In our experiments novices were not familiar with the formal definitions for major and minor excerpts; however, they were familiar with the emotional valence of terms such as happy and sad and used this knowledge, along with corrective feedback, to more appropriately label excerpts on the basis of mode.

Our findings also align with previous research on adults’ and children’s use of statistical regularities in implicit language learning (Reber, 1989; Saffran, Aslin, & Newport, 1996), infants’ preference for culturally familiar rhythms (Hannon & Trehub, 2005b), and learning of musical timbre sequences (Tillmann & McAdams, 2004). Though our novice listeners were not initially able to explicitly relate the major and minor labels to the appropriate excerpts, they did have some implicit understanding of the mode distinction because they rated the major pieces of music as more positive and the minor pieces of music as more negative.

Though we were primarily interested in tracking the process of change in learning the major–minor distinction, we believe our account of the development of this distinction has a number of implications for formal instruction in music, and we also note some possibilities for future research in this area. First, the finding that AV drove learning suggests the efficacy of some popular approaches to music education. Despite the fact that formal definitions (Sadie,
2001) of the major–minor distinction emphasize the whole and half steps that comprise major versus minor scales, music educators have long appreciated the value of providing students with definitions referencing affective quality (happy/sad). Our experiments have shown that music educators are onto something important. Formal definitions referencing musical structure were not particularly helpful for learning to label excerpts by mode, whereas definitions referencing AV led to increased learning, especially among 10-year-olds and adults. Possibly this training would also be effective with younger children given greater preparation prior to training, such as encountering lessons that highlighted affective qualities distinguishing major from minor modes. Future research could examine the effectiveness of such lessons that contrast leitmotifs of cartoon heroes played in the major mode with leitmotifs of villains played in the minor mode.

Further, although providing definitions referencing affective qualities of excerpts do generate gains in accuracy, it appears that individualized feedback on answers is also important, especially when AV alone might lead students to erroneously categorize fast + minor excerpts as major or slow + major excerpts as minor. The benefits of providing individualized feedback are well-known (Siegler, 1996), but the need for such feedback appears especially acute in domains where characteristic errors occur during the course of learning.

Acknowledgments

We would like to thank Mari Jones for her help in composing the musical stimuli.

Notes

1. All participants were given formal, dictionary definitions for major and minor scales. These definitions referred to the unique pattern of whole steps and half steps that sufficiently characterize major and harmonic minor scales. Each participant was told:

   Half-steps appear in the major scale between the third and fourth notes and the seventh and eighth notes. The distance between all other consecutive notes is two half steps, and half steps appear in the minor scale between the second and third notes, the fifth and sixth notes, and the seventh and eighth notes. The distance between all other consecutive notes is two half-steps. The exception is that there are three half-steps between the sixth and seventh notes.

In a pilot study we examined whether these formal definitions improved learning. Surprisingly they did not; rather, formal definitions may have even hindered participants’ attempts to learn to label musical excerpts by mode. In our pilot study of 64 adults, adults who were given both feedback and formal definitions of the major/minor distinction gained only 9% accuracy from pre-test to post-test, whereas adults who were given feedback and no formal definitions showed a 12% increase in categorization accuracy. Throughout this manuscript our reference to receiving definitions therefore refers only to whether subjects received definitions that referenced affective valence.

References


**Biographies**

**Clarissa Thompson** is a developmental psychologist and Assistant Professor of Psychology at the University of Oklahoma. Her research focuses on cognitive development, particularly the way children think, learn, and remember. She has investigated representational change, strategy choice in decision making, and speed of information processing in the domain of number, and her results have implications for mathematics education.

**John Opfer** is a developmental psychologist and Associate Professor at The Ohio State University. His research focuses on mechanisms of cognitive development, particularly development of concepts and reasoning. He has examined how conceptual change occurs in children’s thinking about number and biology, as well as the impact of conceptual change in numerical cognition on economic decision-making.