Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 6, 1417–1432

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.6.1417

Implicit Learning of Nonlocal Musical Rules: Implicitly Learning More Than Chunks

Gustav Kuhn and Zoltán Dienes
University of Sussex

Dominant theories of implicit learning assume that implicit learning merely involves the learning of chunks of adjacent elements in a sequence. In the experiments presented here, participants implicitly learned a nonlocal rule, thus suggesting that implicit learning can go beyond the learning of chunks. Participants were exposed to a set of musical tunes that were all generated using a diatonic inversion. In the subsequent test phase, participants either classified test tunes as obeying a rule (direct test) or rated their liking for the tunes (indirect test). Both the direct and indirect tests were sensitive to knowledge of chunks. However, only the indirect test was sensitive to knowledge of the inversion rule. Furthermore, the indirect test was overall significantly more sensitive than the direct test, thus suggesting that knowledge of the inversion rule was below an objective threshold of awareness.

Keywords: implicit learning, nonlocal rules, inversion rule, mere exposure effect, artificial grammar learning

People often learn about rules and regularities solely through being exposed to stimuli that follow a particular structure, and they apply this knowledge with little or no conscious awareness (Cleeremans, Destrebecqz, & Boyer, 1998). This form of learning is known as implicit learning and is thought to play a major role in different areas of human cognition, such as language acquisition (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997), social context (Lewicki, 1986), and the perception of music (Bigand, Perruchet, & Boyer, 1998; Dienes & Longuet-Higgins, 2004 Tillmann, Bharucha, & Bigand, 2000). Although implicit learning has been investigated using a wide range of paradigms, the artificial grammar-learning task has been the most frequently used task. In a typical artificial grammar-learning task, participants are asked to memorize a set of letter strings that have been created by a complex set of rules. After the memorization phase, participants are then presented with a new set of letter strings that contain grammatical and ungrammatical items and are asked to classify them according to whether they follow the rules. Although participants are usually unable to describe the type of rule they use to classify the test item (Reber, 1989) and often claim that their responses are guesses (Dienes, Altmann, Kwan, & Goode, 1995), they are able to discriminate between grammatical and ungrammatical items. One of the key questions in implicit learning has focused on the way in which the grammar is represented and the type of mechanism or models that could account for this type of learning.

Reber (1967) initially claimed that participants' knowledge of the grammar takes the form of abstract rules that are independent and distinct from the encoding episode. Instance-based models of implicit learning, on the other hand, suggest that the knowledge acquired in artificial grammar learning takes the form of relatively unprocessed learning items, in the form of whole exemplars of the training items (Brooks, 1978; Brooks & Vokey, 1991; Vokey & Brooks, 1992) or parts of the exemplars (i.e., chunks; e.g., Dulany, Carlson, & Dewey, 1984; Perruchet & Pacteau, 1990; Servan-Schreiber & Anderson, 1990), rather than abstract rules. Of these instance-based models, fragment models have become the most popular. Fragment models suggest that learning involves the acquisition of knowledge about the co-occurrences of adjacent letters, also known as chunks. Initial evidence supporting this approach came from a study by Dulany et al. (1984), in which participants were asked to memorize grammatical letter strings, after which they were presented with a set of grammatical and ungrammatical letter strings. In addition to their classification responses, participants were asked to underline the part of the letter string they thought made it grammatical or ungrammatical. These results showed that participants' classification responses were fully accounted for by their knowledge of the permissible letter chunks, thus suggesting that participants merely learned about the co-occurrence of adjacent letters rather than acquired a more abstract representation of the finite-state grammar. Subsequently, several studies have also shown that people learn letter string chunks (e.g., Dienes, Broadbent, & Berry, 1991; Mathews et al., 1989; Perruchet & Pacteau, 1990).

Further support for the fragmentary approach has come from studies in which participants are trained on fragments of the training items rather than the entire items. If participants' knowledge merely takes the form of fragments, there should be no difference in the classification performance between participants trained on the letter string fragments compared with those trained on the entire letter strings. Perruchet and Pacteau (1990) showed that grammaticality judgments for participants who studied grammatical letter strings differed only a small degree from those learning from a list of bigrams making up these letter strings (see

Gustav Kuhn and Zoltán Dienes, Department of Psychology, University of Sussex, United Kingdom.

Correspondence concerning this article should be addressed to Gustav Kuhn, Department of Psychology, University of Durham, South Road, Durham DH1 3LU, United Kingdom. E-mail: gustav.kuhn@durham.ac.uk

also Manza & Reber, 1997, Exp. 5). Few authors would reject the view that chunks play an important role in artificial grammar learning. The question we must ask is whether chunks of adjacent elements are sufficient to account for all implicit learning, as suggested by several popular computational models of implicit learning (Boucher & Dienes, 2003; Perruchet & Vinter, 1998; Servan-Schreiber & Anderson, 1990). For example, the competitive chunker (E. Servan-Schreiber & Anderson, 1990) postulates that learning is a form of chunking mechanism, in which grammaticality judgments are based on a hierarchical network of chunks that, by virtue of having been created from grammatical strings, implicitly encode grammatical constraints. Once exposed to the training items categorization, decisions are made on the basis of the chunk strength. Knowlton and Squire (1994) have formalized the chunk strength in a statistic they call "associative chunk strength" (ACS), which is defined as the frequency with which a chunk occurred in the training set, with the frequency scores averaged across all different chunks. Test items with a high ACS should, therefore, be classified as grammatical and items with lower ACS as ungrammatical.

Several studies have looked at the way different types of knowledge influence people's classification responses, by generating test items in which the test strings' adherence to the finite-state grammar and their association in terms of ACS is manipulated independently. Even though ACS has been shown to account for much of participants' classification performance (Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997), these studies also found independent effects of both rule knowledge[1] and ACS. In Meulemans and Van der Linden's study (1997), great care was taken in balancing both grammatical and ungrammatical test items in terms of global and anchor ACS. Anchor ACS refers to a measure that calculates the frequency with which bigrams and trigrams occur in the beginning and terminal positions. This additional measure was introduced after several studies had shown that participants are particularity sensitive toward initial and terminal bigrams and trigrams (e.g., Reber, 1989). After appropriate training, participants still classified above chance, implying that their knowledge could not be accounted for by bigram and trigram knowledge alone. These results were interpreted as showing that participants acquired knowledge about the abstract structure of the finite state grammar rather than merely knowledge about chunks. However, in a follow-up study, Johnstone and Shanks (1999) demonstrated that, even though the material used was balanced in terms of ACS, information about the legal position of trigrams was ignored. Furthermore, a regression analysis, in which different string familiarities were used to predict participants' classification responses, revealed chunk familiarity rather than grammaticality as a reliable predictor.

In the sort of finite-state grammars that have typically been used in the artificial grammar-learning literature, the regularities of bigrams and trigrams are unavoidably closely linked to the actual finite-state grammar, making it very difficult to isolate the contributions of both rule and chunk knowledge. As shown by Johnstone and Shanks (1999), even if great care is taken in designing material that is perfectly balanced in terms of some measure of ACS, closer inspection of the material is likely to lead to further *n*-gram measures in which the material is not balanced. Although ACS appears to be a reasonable statistical measure of association, we cannot assume that this measure directly reflects the computational processes in the human mind. For example, the letter string fragment *TV* will be noticed more easily than others, which suggests that this fragment will have a disproportionate influence on participants' discrimination performance. We, therefore, cannot assume that all fragments are equally encoded. Furthermore, to date there are several chunking models of implicit learning that vary in the exact computational process used (Boucher & Dienes, 2003; Perruchet & Vinter, 1998; E. Servan-Schreiber & Anderson, 1990). Moreover, the performance of each of these models is also influenced by the parameter values used (see Boucher & Dienes, 2003). Given these variations, it should become apparent that merely balancing the test material in terms of ACS is not a fully satisfactory approach to evaluate whether the knowledge acquired in artificial grammar-learning tasks can be fully accounted for in terms of chunks. Although the material may be balanced in terms of ACS, other potential chunk statistics could be used for correct discrimination.

Chunking models are very good at learning local dependencies but cannot learn nonlocal dependencies, such as the initial *A* in the letter string *AXXB* predicting the *B* in the fourth position, if the intervening material has not been encountered. One way of demonstrating that chunking models cannot account for all implicit learning would be to illustrate that people can implicitly learn nonlocal dependencies that are independent of the chunks. One such rule is the biconditional grammar designed by Mathews et al. (1989), in which rules determine the relationships between nonneighboring letters in a letter string. Several studies have shown that, although people can learn the nonlocal dependency under intentional learning instructions, they failed to learn it under incidental learning conditions (Johnstone & Shanks, 1999; Mathews et al., 1989; Shanks, Johnstone, & Staggs, 1997). Furthermore, Johnstone and Shanks (2001) demonstrated that participants in incidental learning conditions merely learned about chunks.

However, there is evidence suggesting that, under certain conditions, people can learn nonlocal dependencies. Investigations in artificial language learning, in which participants are exposed to streams of syllables or tones that followed a nonlocal rule, have shown that, although people usually learn local dependencies more easily, under certain conditions nonlocal dependencies are learned more readily than local dependencies (Creel, Newport, & Aslin, 2004; Gomez, 2002; Newport & Aslin, 2004). However, in these studies, the role of awareness was never directly assessed, therefore leaving open the questions as to whether knowledge was explicit or implicit.

Artificial grammar learning has been predominantly investigated in the visual modality, and relatively few studies have directly examined implicit learning of music (Altmann, Dienes, & Goode, 1995; Bigand et al., 1998; Dienes & Longuet-Higgins, 2004; Kuhn & Dienes, in press). This seems rather surprising, because learning to perceive structures in music prima facie seems sometimes to occur implicitly. Any nonmusician can spend his or her entire life appreciating music without ever having to explicitly learn about music theory. However, even though he or she may be unaware of the musical grammar, some knowledge about a musical grammar is essential for the aesthetic appreciation of music (Smith & Witt, 1989). Furthermore, many aspects of these grammars are

[1] Rule knowledge referred to knowledge about the finite-state grammar that was independent of the bigram and trigram structure.

learned through a process known as acculturation (Frances, 1988; Krumhansl et al., 2000; Tillmann et al., 2000). With regard to the questions addressed here, the way in which musical grammars are learned implicitly is of particular interest, because many of these grammars are based on nonlocal dependencies and in many cases take the form of algebraic rules, or what Marcus (2001) refers to as operations over variables.

One such grammar underlies 12-tone music, or serialism, a compositional technique introduced by Arnold Schoenberg in the 1920s (see Schoenberg, 1941). According to this method, all 12 pitch classes are placed in a particular order. This tone series forms the basis of the composition, which is transformed using transpose, inversion, retrograde, and retrograde inversion. The aesthetic appreciation of serialist structure depends on (consciously or unconsciously) recognizing the transformations of the original tone series. If people can learn these rules implicitly, it would imply that they had learned a rule that cannot be accounted for by chunking models of implicit learning. Dienes and Longuet-Higgins (2004) showed that people with a special interest in serialist music could implicitly learn serialist transformations. However, because both the training and the test items had several chunks in common, a chunking mechanism could potentially learn to discriminate this type of material. Nonetheless, using a regression analysis, it was shown that participants' knowledge about the transformations was independent of chunks, thus providing tentative evidence that implicit learning of musical structures can go beyond learning chunks.

Kuhn and Dienes (in press) investigated the differences between incidental and intentional learning of a musical rule, a diatonic inversion. An inversion turns the intervals of an original tone series upside down, thus changing its contour without changing the magnitude of the intervals. Figure 1 shows an example of one of the tunes used. All tone series consisted of eight notes, selected from the C-major scale (C_3, D_3, E_3, F_3, G_3, A_3, B_3, C_4).[2] The first four notes could be picked pseudorandomly, and the last four formed the inversion. From Figure 1 it can be seen that the inversion has the same number of diatonic steps but in opposite direction. The advantage of using this type of rule over the normal finite-state grammars is that the rule can be manipulated independently from the bigrams. This rule is a type of biconditional

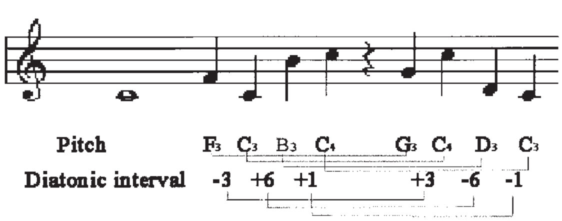

Figure 1. Example of a grammatical training tune. The tune is represented in terms of the diatonic and the chromatic intervals and pitch. Grammatical and ungrammatical tunes differed by two diatonic intervals and two pitches. However, these tunes were compared with each other with regard to associative interval strength and associative pitch strength, which were calculated as described in Experiment 1. There was no significant difference in first-order frequency of intervals between the grammatical (M = 53.0, SD = 8.74) and the ungrammatical (M = 52.7, SD = 6.39), $t(42) < 1$, and no significant difference in associative pitch strength for grammatical (M = 126.5, SD = 11.9) and ungrammatical (M = 126.4, SD = 105), $t(42) < 1$.

grammar because each of the second set of tones or intervals is determined by a mapping from a corresponding tone or interval from the first set. Kuhn and Dienes showed that, although this inversion rule could be learned intentionally, incidental learning merely led to learning chunks. The incidental learning condition used in these experiments was analogous to the incidental learning conditions in the biconditional grammar-learning experiments (Johnstone & Shanks, 2001; Mathews et al., 1989; Shanks et al., 1997). In the following experiments, a similar inversion rule was used. However, participants were exposed to a larger repertoire of training tunes, which may enhance rule learning (cf. Bishop, 1996). The aim of the experiments was to establish the type of features participants could learn through exposure to grammatical tunes by assessing their knowledge on different sets of test tunes. These test sets were designed by manipulating the type of knowledge that was required to distinguish between grammatical and ungrammatical tunes. In particular, the tunes in one of the test sets was created from bigrams that never occurred in the training set. Any learning mechanism that is solely based on learning chunks of adjacent elements would, therefore, fail to deal with this type of material. Comparing participants' performances on these different test sets would then reveal the type of knowledge they acquired.

One of the major problems in the implicit learning literature is how to measure implicit or unconscious knowledge. Knowledge of some content is unconscious if one knows the content but is not conscious of knowing it (Rosenthal, 2002). If one were not conscious of knowing some content, one would not verbally report using it. Thus, participants' failure to report the grammar has been used to justify that the knowledge acquired was implicit (Reber, 1989). However, the use of verbal recall to assess participants' explicit knowledge has been criticized by many authors on the grounds that it is an insensitive and incomplete measure of participants' awareness (see Berry & Dienes, 1993 Dulany et al., 1984; Shanks & St. John, 1994). Subsequently, several more stringent criteria of awareness have been proposed. Subjective measures of the conscious status of knowledge states test whether participants are conscious of being in knowledge states by directly asking them to report what mental state they are in. Two such criteria that can be used: the guessing and the zero-correlation criteria (see Dienes, 2004; Dienes & Perner, 2004). According to the guessing criterion, knowledge is implicit if participants perform above chance when they believe they are guessing (Cheesman & Merikle, 1984; Dienes et al., 1995). The guessing criterion implies that there is implicit knowledge without ruling out the possibility that there is explicit knowledge on the nonguessing trials. The explicit component of people's knowledge can be measured using the zero-correlation criterion. For the zero-correlation criterion, participants' classification performance is plotted against confidence ratings, which are obtained after each response. If participants' classification performance improves with increasing confidence (positive slope), we can claim that participants are aware of the epistemic status of their mental states and hence have metaknowledge about their knowledge. However, if there is no such relationship, participants do not know when they know and when they are guessing, and hence their knowledge is implicit (see also Dienes & Altmann, 1997; Dienes et al., 1995; Dienes & Longuet-Higgins,

[2] C_3 refers to the middle C and C_4 to the C one octave above.

2004; Kelley, Burton, Kato, & Akamatsu, 2001; Newell & Bright, 2002; Tunney & Shanks, 2003).

Another approach uses the following logic. If I am conscious of knowing some content, I will certainly be able to express that content when directly asked for it. If I am not conscious of knowing some content, it may or may not be elicited when directly asked. However, it must be used in some way to count as knowledge. Presumably, it will be used to perform the task for which the system acquiring the content was adapted. In sum, unconscious knowledge might express itself more sensitively on an indirect rather than a direct test but conscious knowledge will always produce performance on a direct test at least as good as that on an indirect test (e.g., Reingold & Merikle, 1988, 1993; Shanks & St. John, 1994). Finding better performance on an indirect rather than a direct test is known as the objective threshold criterion of unconscious knowledge; direct tests are tasks in which participants are explicitly instructed to discriminate stimuli according to the distinction in question (e.g., grammatical vs. ungrammatical), and indirect tests make no reference to the distinction in question (e.g., possibly liking ratings; Zajonc, 1968).

It is difficult to predict in advance whether unconscious knowledge will express itself on a direct test. If knowledge is expressed on a direct test, one can test its conscious or unconscious status by the use of subjective measures, like the guessing and zero-correlation criteria. If knowledge does not express itself on a direct test, it still might express itself on an indirect test specially chosen to be the sort of test relevant to the expression of that sort of knowledge in ecological contexts.

Numerous studies have shown that unreinforced exposure to a stimulus leads to an increase in positive attitude toward that stimulus (e.g., Bornstein, 1989; Zajonc, 1968). Similarly, it has been shown that if participants have knowledge about structures, this leads to an increase in liking for items that have the same structure (Gordon & Holyoak, 1983; Manza & Reber, 1997; Newell & Bright, 2003; Whittlesea & Price, 2001). Because participants are merely asked to rate items according to how much they are liked, no reference is made to the grammaticality distinction; thus, liking ratings fulfill the criterion of being an indirect test. Further, the reaction that people habitually produce to music is to both feel and express how much they like it. If knowledge of musical structures (unlike letter string structures) will express itself in any way, it plausibly will express itself in terms of an aesthetic judgment (e.g., Meyer, 1903; Peretz, Gaudreau, & Bonnel, 1998; Verveer, Barry, & Bousfield, 1933; Wilson, 1979). In the current experiments, an increase in positive affect for grammatical as opposed to ungrammatical items, resulting from exposure to grammatical training tunes, was used as evidence for the existence of knowledge about the training tunes. Participants' awareness of this knowledge was then measured using a direct test, in which they were asked to make classification responses followed by confidence ratings. Knowledge was claimed to be unconscious if either of two criteria were met. First, participants in the experimental group performed no better than chance, or the control group, and the indirect test was more sensitive than the direct one (Reingold & Merikle, 1988); this is sometimes called the objective threshold of conscious awareness (Cheesman & Merikle, 1984). Second, if performance on the direct test was above chance, confidence ratings were used to measure participants' awareness of having knowledge (sometimes called the subjective threshold of awareness; Cheesman & Merikle, 1984). Although the two criteria—subjective and objective—are referred as different thresholds, they are simply different methodological ways of measuring whether a person has knowledge that they are not aware of having.

Experiment 1A

The aim of Experiment 1A was to establish whether participants would become sensitive toward the regularities in the training tunes by measuring their knowledge on an indirect test. After listening to a series of grammatical tunes (i.e., tunes instantiating an inverse), participants were asked to rate how much they liked the new set of tunes. Any effect of grammaticality on participants' liking ratings would indicate that they had acquired some form of knowledge about the structure of the training tunes. Three different test sets were designed singling out different cues that could be used to discriminate between grammatical and ungrammatical tunes. In the first set (exemplar), grammatical tunes were taken from the training set, whereas all ungrammatical tunes were novel and consisted of bigrams, which rarely occurred in the training set (low ACS). Any sensitivity toward the grammaticality in this set could result from knowledge about the unprocessed training items, knowledge about chunks, or knowledge about the inversion rule. In the second set (fragment), all test tunes were novel. However, the grammatical tunes were generated from a set of *n*-grams, which occurred more often in the training phase than did the *n*-grams for the ungrammatical tunes. Because all the tunes used in this set were novel, any sensitivity toward the grammaticality could no longer be attributed to knowledge about whole items and must, therefore, be due to knowledge about the ACS or knowledge about the inversion rule. The final set (abstract) was designed to establish whether learning could go beyond the learning of adjacent elements. Both grammatical and ungrammatical tunes were generated from a novel set of bigrams, none of which ever occurred in the training phase. Any sensitivity toward the grammaticality could no longer be due to knowledge about fragments and would indicate that participants' knowledge went beyond the learning of chunks.

Because of the nature of the material, it is possible that certain tunes sounded more pleasing than others regardless of the training. This problem was circumvented by comparing the experimental group's ratings with those of a control group, whose procedure only differed in that they were not exposed to the training tunes.

Method

Participants. Ninety-six individuals from the University of Sussex took part in the experiment. Participants in the experimental group were paid £4, and those in the control group were paid £3. None of the participants had previously taken part in any implicit learning experiments. Participants were randomly allocated to either the experimental or control group, with an equal number allocated to each group.[3]

Materials. One of the most salient features of music is that of pitch, because it allows for the creation of melodies. The perception of music

[3] There was no difference in musical experience between participants in the control group and the experimental group, $\chi^2 = 0.67$, $p = .27$. Musical experience was defined in terms of whether participants had more than 3 years of formal music education.

Table 1
Mean ACS and MFF for Grammatical and Ungrammatical Tunes for the Exemplar and the Fragment Set in Terms of Diatonic Intervals, Pitches, and Chromatic Intervals

	Pitch				Diatonic interval				Chromatic interval			
	G		UG		G		UG		G		UG	
Test set and statistic	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Exemplar MFF	118.90	13.40	116.10	7.87					49.13	3.69	49.66	5.59
Exemplar global ACS	11.37[a]	3.46	8.92	2.75	11.19[a]	0.87	3.34	1.77	5.85[a]	1.06	1.78	1.39
Exemplar anchor ACS	2.17[a]	0.53	1.33	0.60	2.35[a]	0.53	0.42	0.40	2.52[a]	0.72	3.87	0.77
Fragment MFF	125.80	12.60	120.70	6.39					51.70	8.46	50.50	7.76
Fragment global ACS	14.13[a]	3.46	9.12	3.01	8.81[a]	1.61	3.09	1.44	4.86[a]	1.24	1.73	0.82
Fragment anchor ACS	2.29[a]	0.80	1.44	0.83	1.23[a]	0.69	0.40	0.43	6.58[a]	0.96	8.62	0.71

Note. G = grammatical; UG = ungrammatical; MFF = mean feature frequency; ACS = associative chunk strength.
[a] Significant difference between the grammatical and the ungrammatical items, $p < .05$.

relies on many different perceptual dimensions, such as timbre, loudness, rhythm, and pitch, all of which are combined to form a particular piece of music. However, for the purpose of this article, only the pitch dimension is relevant, because grammaticality is defined in terms of the tune's pitches, and all other factors are held constant. When memorizing a tune, people do not merely represent the melody as a series of independent pitches. They tend to become sensitive toward the distance between two notes, which can be measured by the chromatic interval (see Cross, 1997; Krumhansl, 1991; Longuet-Higgins, 1987; Shepard, 1982). The pitch distance between one note and the next nearest note is called a semitone. The distance between a C and a G is, therefore, seven semitones. A melody can thus be represented as both a sequence of pitches and a sequence of chromatic intervals. Our experience of music is also influenced by musical schemas, such as tonal scales. These schemas guide us as to what aspect of the melody we remember. In Western music there are two main scales: major and minor. In the C-major scale, all notes are one tone apart (one tone = two semitones) except for E–F and B–C, which are one semitone apart. The C-major scale, therefore, consists of seven different notes, and the distance between two notes is called a diatonic interval. The distance between C and G would therefore be +4. A further feature that is of particular importance is the contour, or the pattern of ups (+) and downs (–) of pitches from one note to the other (see Dowling & Harwood, 1986).

The material was designed so that chunking statistics were manipulated independently from the actual structural rule, which in this case was characterized by the diatonic inversion. The rule referred to the actual inversion rules that constituted the grammar to be learned, whereas fragment knowledge was defined in terms of pitches, diatonic and chromatic interval, and contour. Because simple melodies can be represented in so many different ways, great care was taken to control all the just-mentioned stimulus dimensions.

The grammar used was an inversion rule, similar to that used by Kuhn and Dienes (in press). All tunes consisted of eight notes, which were selected from the C-major scale. These notes can be numbered from 1 to 8 (pitch number): $C_3 = 1$; $D_3 = 2$; $E_3 = 3$; $F_3 = 4$; $G_3 = 5$; $A_3 = 6$; $B_3 = 7$; $C_4 = 8$, where C_3 is middle C. The first four notes formed the prime and were selected semirandomly, whereas the last four notes formed the inversion, which was created by subtracting the pitch number from a constant (9). The prime 3 6 4 3 leads to the following inversion 6 3 5 6, and the tune 3 6 4 3–6 3 5 6. In terms of diatonic intervals, the tune is defined as + 3–2–1 + 3 −3 + 2 + 1.

We constructed 120 different grammatical training tunes. These tunes were created from a unique set of interval bigrams, ensuring that a new set with different interval bigrams could be designed.[4] Care was taken to balance the contour patterns. The following contour patterns each occurred 20 times as a prime ++–, –+, +–+, –+-, +–, and –+.

Three different sets of test tunes were created, which differed in the way they were associated with the training set. For the exemplar set, 12 tunes were selected from the training set, which formed the grammatical tunes. Twelve ungrammatical tunes were created that violated the inversion rule in both the interval magnitude and contour. Care was taken to ensure that grammaticality was not correlated with first-order frequency of the diatonic intervals. Diatonic intervals occurred the same number of times in each position (e.g., the interval +4 occurred twice in Position 2 for both grammatical and ungrammatical items). Furthermore, a new statistic called mean feature frequency (MFF) was created to ensure that grammatical and ungrammatical tunes were balanced in terms of pitch and chromatic interval first-order frequency. MFF was calculated for each tune by averaging the number of times each pitch class (or chromatic interval) occurred in the training set in each of the eight positions. Table 1 shows the MFF for grammatical and ungrammatical tunes in terms of pitches and chromatic intervals. From Table 1 it can be seen that there was no significant difference in MFF between grammatical and ungrammatical tunes. There was no need to calculate MFF in terms of diatonic intervals, because grammatical and ungrammatical tunes were created from identical diatonic intervals. The ungrammatical tunes were created by using interval combinations (bigrams) that rarely, or never, occurred in the training set. Table 1 shows ACS statistics for diatonic intervals, pitch classes, and chromatic intervals. It can be seen that, apart from chromatic interval anchor ACS, grammatical items had significantly higher ACS statistics than the ungrammatical items. Care was taken that both grammatical and ungrammatical tunes had exactly the same number of contour bi- and trigrams (adjacent contours, e.g., ++, +–+)

For the fragment set, 12 novel grammatical tunes were created with both high global and anchor ACS. The 12 ungrammatical tunes were created using the same rationale as in the exemplar set. Diatonic intervals occurred the same number of times in each position as the grammatical items. Table 1 shows that there was no significant difference in MFF statistics either in terms of pitches and chromatic intervals. However, apart from the chromatic interval anchor ACS, the ACS statistics for all other measures were significantly higher in the grammatical compared with the ungrammatical items.

For the abstract set, both grammatical and ungrammatical tunes were created from a novel set of interval bigrams,[5] which never occurred during the training phase. This meant that none of the tunes had any bigrams in

[4] See http://www.lifesci.sussex.ac.uk/home/Gustav_Kuhn/Kuhn_DienesJEP_LMC2006/index.htm

[5] Using novel interval bigrams meant that pitch bigrams were novel too.

Table 2
Mean Liking Ratings Given for Grammatical and Ungrammatical Tunes for Each of the Three Test Sets

	Exemplar				Fragment				Abstract			
	G		UG		G		UG		G		UG	
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experimental	5.28	0.83	4.96	0.65	5.44	1.09	4.94	0.98	5.40	0.70	5.09	0.86
Control	5.20	0.81	5.13	0.79	4.96	0.81	5.10	0.86	5.39	0.76	5.63	0.58

Note. G = grammatical; UG = ungrammatical.

common with the training set, thus leading to zero ACS. This also applied to pitch and chromatic interval representations. The ungrammatical tunes violated the inversion rule both in their interval magnitude and contour. To ensure that they did not differ in interval or pitch class first-order frequencies, the ungrammatical tunes were created by using the same prime and inversions as the grammatical tunes but changing the way in which they were combined. Furthermore, they were combined in a way that ensured that grammatical and ungrammatical tunes had the same number of contour bi- and trigrams.[6]

The music notes (C_3, D_3, E_3, F_3, G_3, A_3, H_3, C_4) were sampled (22.5 kHz) using a Yamaha P50 Sound Box (grand piano). The starting note lasted 1,200 ms and remained the same for each tune (C_3). The duration of the other notes was 300 ms, except for the fourth and eighth tones, which lasted 600 ms. This meant that the tunes were perceived as having a gap between the prime and the inversion. The tunes were produced by concatenating these individual samples.

Application. The tunes were presented over a pair of headphones, using a Power Mac. A computer program was written that presented the tunes in a different random order for all participants and recorded their keyboard responses.

Procedure. The experiment consisted of two parts: a training phase and a test phase. Participants in the experimental group were informed that they were taking part in a memory experiment. They were presented with 120 grammatical tunes and asked to memorize them as far as possible. After each tune they were asked to indicate whether they thought the tune had been played before by pressing the appropriate key. After the training phase, participants were instructed that they were about to hear a new set of tunes similar to the tunes they heard before. Participants were asked to rate how much they liked them on a scale ranging from 1 to 9 (1 = *do not like it,* 5 = *indifferent,* 9 = *like it a lot*). They were also encouraged to make use of the full range of the scale. Participants in the control group took part in the same test sets and were given the same instructions as the experimental group but did not take part in the training phase.

Results

Table 2 shows participants' mean liking ratings of both grammatical and ungrammatical tunes for each of the three sets. A three-way analysis of variance (ANOVA) on liking ratings, with group (experimental vs. control) and test set (exemplar vs. fragment vs. abstract) as between-subjects variables and grammaticality (grammatical vs. ungrammatical) as a within-subjects variable, found no significant effect of grammaticality, $F(1, 90) = 2.47$, $MSE = 0.67$, $p = .12$, no significant effect of group, $F(1, 90) = 0.045$, $MSE = 0.047$, $p = .83$, no significant effect of test set, $F(2, 90) = 2.89$, $MSE = 1.45$, $p = .26$, and no significant Group × Test Set × Grammaticality interaction, $F(2, 90) = 0.33$, $MSE = 0.089$, $p = .72$. However, there was a significant Group × Grammaticality interaction, $F(1, 90) = 11.68$, $MSE = 3.17$, $p = .001$, whereby the experimental group rated grammatical tunes more highly than ungrammatical tunes, $t(47) = 3.61$, $p = .001$, but there was no significant difference for the control group, $t(47) = -1.32$, $p = .20$. Furthermore, planned post hoc analysis showed that in the abstract set the experimental group rated grammatical tunes as more likable than ungrammatical tunes, $t(47) = 3.61$, $p = .001$, whereas no such difference was found in the control group, $t(47) = -1.32$, $p = .2$.

Discussion

The results from Experiment 1A showed that exposure to the training tunes led to an increase in liking for grammatical tunes versus ungrammatical tunes. Participants in the experimental group were, therefore, able to discriminate between grammatical and ungrammatical tunes using liking ratings. Furthermore, the nonsignificant Group × Test Set × Grammaticality interaction showed that this effect was independent of test set. The fact that there was no difference in mere exposure effect between these three test sets implies that liking ratings were unaffected by ACS or whether the tunes were identical to the training set. Furthermore, planned post hoc analysis showed that, in the abstract set, the experimental group rated grammatical tunes as more likable than ungrammatical tunes; no such difference was found in the control group. These results imply that participants were able to discriminate between grammatical and ungrammatical tunes even when all test tunes were created from novel bigrams. However, before concluding whether participants' discrimination abilities were due to implicit or explicit knowledge, participants' awareness of this knowledge must be assessed.

Experiment 1B

Experiment 1A showed that participants could discriminate between grammatical and ungrammatical tunes using liking ratings. The aim of Experiment 1B was to establish participants' levels of awareness of this knowledge. Most studies that have investigated the relationship between affective and classification judgments have used memory rather than learning paradigms. Newell and Bright (2001) used an artificial grammar-learning task in which

[6] A full list of the material can be found on http://www.lifesci.sussex.ac.uk/home/Gustav_Kuhn/Kuhn_DienesJEP_LMC2006/index.htm

Table 3
Mean Classification Performances (Percentage of Correct Responses) for Each of the Three Test Sets

	Exemplar				Fragment				Abstract			
Group	*M*	*SD*	LCI	UCL	*M*	*SD*	LCI	UCL	*M*	*SD*	LCI	UCL
Experimental	56.8	10.2	51.3	62.2	58.9	8.7	54.2	63.5	51.6	10.1	46.2	56.9
Control	50.8	6.7	47.2	54.3	50.8	11.4	44.7	56.9	51.6	11.3	45.6	57.6

Note. LCI = lower 95% confidence interval; UCL = upper 95% confidence interval.

participants' knowledge was assessed using both liking and rule judgments. They showed that liking ratings were sensitive to grammaticality (cf. Gordon & Holyoak, 1983; Manza & Bornstein, 1995). They also consistently showed rule judgments to be more sensitive than liking ratings. However, Newell and Bright used letter strings rather than melodies. In addition, whereas the everyday function of music relates crucially to its aesthetic appeal, this is not true of letter strings. Thus, it is possible that liking may be a more sensitive measure of the learning of musical structures than rule judgments.

The relationship between the mere exposure effect and individuals' awareness of the source of this effect is a matter of debate. According to the misattribution theory (Bornstein & D'Agostino, 1992, 1994), the mere exposure effect for the repetition of old stimuli arises from interpreting processing fluency as liking rather than attributing it to the fact the item is old. Once the stimuli are recognized, the perceptual fluency can be correctly attributed to previous exposure, thus removing the increase in liking. The mere exposure effect can, therefore, only occur as long as participants are unaware of the source of the perceptual fluency, which is supported by experiments showing a mere exposure effect in the absence of recognition (e.g., Kunst-Wilson & Zajonc, 1980; Seamon, Brody, & Kauff, 1983; Seamon, McKenna, & Binder, 1998). By this account, an increase in liking to grammatical stimuli for musical inverses after training on inverses occurs as long as people are not aware that the grammatical stimuli have the same structure as the training stimuli.

The two-factor model, alternatively, assumes that the increase in liking results from two opposing processes: habituation and arousal (e.g., Berlyne, 1970; Lee, 2001). Unlike the misattribution model, the two-factor model assumes the same knowledge can be responsible for both recognition and liking performance (Lee, 2001). However, this does not necessarily imply that classification performance must be above chance, because the knowledge could be implicit rather than explicit (Bornstein, 1989), thus leading to a mere exposure effect in the absence of recognition.

We will not specifically presume either theory. We merely assume that the liking increase in Experiment 1A resulted from sensitivity to either the inversion rule or a regular consequence of it, and if the knowledge of this regularity is conscious, participants asked to search for rules will use the knowledge in classification.

The material used in Experiment 1B was identical to that in Experiment 1A. However, rather than using liking ratings, participants' knowledge was measured using a direct test. After the training phase, participants were informed about the existence of a rule and asked to classify accordingly. If the knowledge acquired in Experiment 1A was below an objective threshold, then participants in the experimental group should not perform better than chance or better than the control group.

After each classification, participants were also asked to rate how confident they were about their decision. If participants' discrimination performance is above chance, confidence ratings are used to measure metaknowledge. The guessing and the zero-correlation criteria were used to determine the explicit and implicit components of this knowledge (see Dienes, 2004; Dienes & Perner, 2001, 2004; and Dienes & Scott, 2005, for complementary discussions of assumptions of using subjective measures of awareness).

Method

Participants. Ninety-six individuals from the University of Sussex took part in the experiment (48 in the experimental group, 48 in the control group). Participants in the experimental group were paid 4 pounds, whereas those in the control group were paid 3 pounds. None of the participants had previously taken part in any implicit learning experiments. Participants were randomly allocated to either the experimental or the control group.[7]

Material. This was identical to that used in Experiment 1A.

Procedure. The material and procedure for the training phase were identical to those in Experiment 1A. However, after the training phase, participants were informed about the existence of a rule used to generate the pitches of all the tunes heard before. They were then asked to listen to the new set of tunes, half of which were grammatical, and were asked to classify the new set, by pressing the appropriate key, based on whether they thought that these new tunes followed the same pattern or structure as the tunes just memorized. After each classification, participants were asked to rate how confident they felt about their decision, using a confidence rating that ranged from 50% to 100% (50%, 51–60%, 61–70%, 71–80%, 81–90%, 91–100%). Participants were explicitly informed that 50% confidence meant a literal guess. Participants were then presented with one of the three test sets. After the completion of this task, they were asked to write down the strategy they used to classify the tunes. The procedure for participants in the control group was the same as that of the experimental group except that they did not take part in the training task (see Dienes & Altmann, 2003, for a discussion of using an untrained control group).

Results

Classification performance. Table 3 shows the mean classification performance for participants in the experimental and the control groups for each of the three test sets.

A two-way ANOVA, with group (experimental vs. control) and test set (exemplar vs. fragment vs. abstract) as between-subjects

[7] There was no difference in musical experience between participants in the control group and the experimental group, $\chi^2 = 1.51$, $p = .15$.

variables, found a significant effect of group, $F(1, 90) = 5.43$, $MSE = 527.3$, $p = .022$, but no significant effect of test set, $F(2, 90) = 0.91$, $MSE = 88.4$, $p = .41$, and no significant Group × Test Set interaction, $F(2, 90) = 1.45$, $MSE = 140.5$, $p = .24$. Participants in the experimental condition, therefore, performed better than those in the control group. Furthermore, the nonsignificant Group × Test Set interaction implies that the difference between the experimental and the control groups was not different for the various test sets. However, the confidence intervals in Table 3 indicate that only participants in the experimental group performed significantly above chance in the exemplar and fragment sets but failed to do so in the abstract set. Furthermore, planned post hoc tests showed that the experimental group performed significantly better than the control group in both the exemplar set, $t(30) = 1.97$, $p = .029$, one-tailed, and the fragment set, $t(30) = 2.25$, $p = .032$, but there was no such difference in the abstract set, $t(30) < 1$.

Participants' subjective awareness was only analyzed for conditions in which learning was observed. The zero-correlation criterion was analyzed by calculating the gamma correlation coefficient between confidence rating and classification performance for each individual participant and averaging across participants. Although there appears to be a difference between the gamma scores on the exemplar ($M = 0.054$, $SD = 0.48$)[8] and the fragment ($M = 0.34$, $SD = 0.33$) sets, this difference was not quite statistically significant, $F(1, 29) = 3.84$, $MSE = 0.639$, $p = .06$. The gamma scores on the two test sets were, therefore, combined, which revealed that they were significantly greater than zero, $F(1, 29) = 7.27$, $MSE = 1.21$, $p = .01$. According to the zero-correlation criterion, participants' knowledge of the grammatical status of tunes differing according to the chunks in training tunes was, therefore, at least partly explicit. Table 4 shows participants' classification performance when they were guessing (50% confidence rating). An ANOVA with group (experimental vs. control) and set (exemplar vs. fragment) as within-subjects variables on guessing scores showed no significant effect of set, $F(1, 39) = 0.727$, $MSE = 385.3$, $p = .40$, no significant effect of group, $F(1, 39) = 0.096$, $MSE = 50.6$, $p = .75$, and no significant Group × Set interaction, $F(1, 39) = 0.216$, $MSE = 114.4$, $p = 22$. There was, therefore, no evidence to suggest that participants in the experimental group performed significantly better than the control group.

Participants' verbal descriptions of the strategies used to discriminate the test tunes provided no valuable insight into their declarative knowledge.

Table 4
Mean Classification Performance (Percentage of Correct Responses) When Participants Were Guessing (50% Confidence Rating) and the Number of Participants Who Were Included in the Analysis

	Experimental			Control		
Test set	*M*	*SD*	*N*	*M*	*SD*	*N*
Exemplar	56.8	16.6	14	56.2	20.4	16
Fragment	54.0	30.4	13	42.4	23.0	16

Note. Several participants were excluded because they never guessed.

Discussion

The aim of Experiment 1B was to establish whether participants could distinguish between grammatical and ungrammatical tunes when they were explicitly asked to do so. The results showed that participants in the experimental condition performed significantly better than the control group. Post hoc inspection of the results revealed that participants in the experimental group only performed significantly better than the control group (and better than chance) in the exemplar and the fragment sets. In the abstract set, the experimental group performed at chance and no better than the control group. Although the Group × Test Set interaction failed to reach significance, these results could indicate that knowledge about fragments and exemplars was above an objective threshold of awareness, and knowledge used in the abstract set was possibly below an objective threshold of awareness.

In the exemplar and the fragment sets, participants in the experimental group performed significantly better than chance and the respective control group. Participants' awareness was, therefore, analyzed with regard to their subjective level of awareness. Because the experimental group performed no different from chance or the control group on the abstract set, there was no point in analyzing their classification responses. The explicit component of participants' classification performance was assessed using the zero-correlation criterion. These results showed a positive zero-correlation criterion on the exemplar and fragment sets, with no significant difference between the two; thus, at least part of the participants' knowledge was explicit. The implicit component of participants' classification performance was assessed using the guessing scores. This analysis showed that the experimental group performed no different from the control group on either of the test sets when they were guessing, thus providing no evidence to suggest that their knowledge was implicit.

Experiment 2

The aim of Experiment 2 was to replicate the results found in the abstract set in Experiments 1A and 1B and provide evidence that knowledge of the inversion rule was implicit. Experiment 2 differed from Experiment 1 in using just the abstract set, and the same participants were asked to make both rule and liking judgments for the test tunes.

Method

Participants. Ninety-six individuals from the University of Sussex took part in the experiment (48 experimental group, 48 control group). Participants in the experimental group were paid 4 pounds, and those in the control group were paid 3 pounds. None of the participants had previously taken part in any implicit learning experiments. Participants were randomly allocated to either the experimental or the control group.[9]

[8] Data from 1 participant were excluded because this participant gave identical confidence ratings for each response, thus preventing the calculation of a correlation coefficient.

[9] There was no difference in musical experience between participants in the control group and the experimental group, $\chi^2 = 0$, $p = 1$.

Table 5
Mean Liking Ratings Given for Grammatical and Ungrammatical Tunes, and the Mean Classification Performance (Percentage of Correct Responses)

	Liking ratings							
	G		U		Classification performance			
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	LCI	UCL
Experimental	5.06	0.99	4.88	0.94	50.81	7.81	48.54	53.08
Control	5.11	0.76	5.18	0.72	51.91	6.36	50.06	53.76

Note. LCI = lower 95% confidence interval; UCL = upper 95% confidence interval; G = grammatical; U = ungrammatical.

Material. Training material was identical to that in Experiments 1A and 1B. Thirty-six test tunes were created by the same method as the abstract set in Experiments 1A and 1B.[10]

Procedure. The training procedure was identical to that in Experiments 1A and 1B. After the training phase, participants were given the liking test used in Experiment 1A and the classification test from Experiment 1B. The presentation order of the tests was counterbalanced across participants. The procedure for participants in the control group was the same as that of the experimental group except that they did not take part in the training phase.

Results

Liking ratings. Because there were no significant effects of order on the liking ratings, the data were collapsed over this variable. Table 5 shows participants' mean liking ratings for grammatical and ungrammatical tunes.

A two-way ANOVA with group (experimental vs. control) as between-subjects variable and grammaticality (grammatical vs. ungrammatical) as within-subjects variable on liking ratings found no significant effect of group, $F(1, 94) = 1.12$, $MSE = 1.50$, $p = .29$, and no significant effect of grammaticality, $F(1, 94) = 1.13$, $MSE = 0.158$, $p = .29$. However, there was a significant Group $\times$ Grammaticality interaction, $F(1, 94) = 5.65$, $MSE = 0.785$, $p = .019$. Participants in the experimental group rated the grammatical tunes more highly than the ungrammatical tunes, $t(47) = 2.48$, $p = .017$. However, there was no such difference for the control group, $t(47) < 1$.

Classification response. Because there were no significant effects of order on the classification responses, the data were collapsed over this variable. Table 5 shows participants' rule judgment performance. The experimental group did not perform significantly better than the control group, $t(94) = -0.76$, $p = .45$. Surprisingly, the control group performed numerically better than the experimental group. From the confidence intervals, it can be seen that only the control group performed significantly better than chance.

Discussion

Experiment 2 showed that exposure to the training tunes led to an increase in liking ratings for grammatical relative to ungrammatical tunes. Participants in the experimental group were, therefore, able to discriminate between grammatical and ungrammatical tunes using affect ratings. However, when these participants were explicitly asked to discriminate between grammatical and ungrammatical tunes, they did not perform significantly better than a control group nor significantly better than chance. In fact, the experimental group performed numerically lower than the control group, thus providing tentative evidence to suggest that participants' knowledge was below an objective threshold of awareness.

So far we have demonstrated that people could learn an inversion rule that was independent in terms of fragment knowledge, defined in terms of statistical regularities of adjacent diatonic and chromatic intervals and pitch classes. The inversion rule used in this experiment was defined both in terms of the interval magnitude and the interval contour. Although great care was taken to balance the test tunes in terms of contour *n*-grams, this was only partially successful. Grammatical and ungrammatical tunes could be broken down into identical contour bi-, tri-, and tetragrams. However, the grammatical tunes had more higher order dependencies (greater than tetragrams) in common with the training tunes. With regard to contour *n*-grams with and above 5-g, grammatical items ($M = 30.0$, $SD = 2.28$) had a substantially higher contour ACS than the ungrammatical items ($M = 11.5$, $SD = 1.23$), $t(34) = 30.2$, $p < .0005$. Previous work has shown that the contour plays an important role in the recognition of tunes (Dowling, 1971). It, therefore, seems reasonable to suggest that participants were influenced by contour cues. Before any definitive conclusions can be made about whether implicit learning can go beyond the learning of chunks, it is essential to rule out the possibility that participants' discrimination abilities were based on contour 5-g.

Experiment 3

The aim of Experiment 3 was to rule out the possibility that participants' discrimination abilities could be based on contour information. The test material used in the following experiment was created from a novel set of interval bigrams, just as in the previous experiment. However, ungrammatical tunes violated the inversion rule only in terms of interval magnitude rather than contour. Grammatical and ungrammatical tunes, therefore, had the same contour, thus excluding the possibility that this type of knowledge could be used for the correct discrimination. This material was thought to be more challenging than that used in the previous experiments. Furthermore, Kuhn and Dienes (in press)

[10] A full list of the material can be found on http://www.lifesci.sussex.ac.uk/home/Gustav_Kuhn/Kuhn_DienesJEP_LMC2006/index.htm

Table 6
Mean Liking Ratings Given for Grammatical and Ungrammatical Tunes and the Mean Classification Performance (Percentage of Correct Responses)

	Liking ratings							
	G		U		Classification performance			
Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	LCI	UCL
Experimental	5.20	0.75	5.08	0.71	50.47	6.09	48.70	52.24
Control	5.05	0.75	5.15	0.70	49.43	6.14	47.65	51.21

Note. LCI = lower 95% confidence interval; UCL = upper 95% confidence interval; G = grammatical; U = ungrammatical.

showed that discriminating between grammatical and ungrammatical tunes was much harder if the inversion rule is determined purely in terms of the interval magnitude rather than the contour. However, if learning can be observed on this type of material, strategies based on learning adjacent chunks can be ruled out.

Method

Participants. Ninety-six individuals from the University of Sussex took part in the experiment (48 experimental group, 48 control group). Participants in the experimental group were paid 4 pounds, and those in the control group were paid 3 pounds. None of the participants had previously taken part in any implicit learning experiments. Participants were randomly allocated to either the experimental or control group.[11]

Material. Training material was derived from that used in Experiment 2. Twenty-four ungrammatical tunes were created from the same set of interval bigrams as in Experiment 2. The ungrammatical tunes were created by exchanging the last four intervals between two tunes that had the same contour pattern. This meant that grammatical and ungrammatical tunes consisted of novel interval (chromatic and diatonic) and pitch *n*-grams. Furthermore, because both grammatical and ungrammatical tunes had identical contours, contour information could not be used for correct discrimination.[12]

Procedure. This was the same as in Experiment 2

Results

Liking ratings. Preliminary analysis of the liking data revealed a significant preference for ungrammatical tunes by the control group. This was attributed predominantly to one ungrammatical tune ($M = 3.6$, $SD = 2.01$). This tune was removed along with its counterpart tune, which was required to keep the material in balance (one ungrammatical and two grammatical).[13] Because there were no significant effects of order on liking ratings, the data were collapsed over this variable. Table 6 shows mean liking ratings given by the experimental and the control groups for grammatical and ungrammatical tunes. A two-way mixed-model ANOVA, with group (experimental vs. control) as its between-subjects variable and grammaticality (grammatical vs. ungrammatical) as the within-subjects variable on liking ratings, found no significant effect of grammaticality, $F(1, 94) = 0.015$ $MSE = 0.0018$, $p = .90$, and no significant effect of group, $F(1, 94) = 0.068$, $MSE = 0.064$, $p = .80$. However, the was a significant Group × Grammaticality interaction, $F(1, 94) = 4.96$ $MSE = 0.60$, $p = .028$. Participants in the experimental group rated the grammatical tunes more highly than the ungrammatical tunes, $t(47) = 1.72$, $p = .04$, one-tailed. However, there was no such difference for the control group, $t(47) = -1.47$, $p = .15$.

Classification performance. Because there were no significant effects of order on the classification performance, the data were collapsed over this variable. Table 6 shows participants' rule judgment performance. From the confidence intervals, it can be seen that neither group performed significantly better than chance. Furthermore, the experimental group did not perform significantly better than the control group, $t(94) < 1$.

Several studies have shown that participants are sensitive toward patterns of repeating elements, known as repetition structure (Brooks & Vokey, 1991; Gomez, Gerken, & Schvaneveldt, 2000; Mathews & Roussel, 1997). The effect of repetition structure on participants' liking ratings was investigated using a regression analysis. For each test tune the number of times its repetition structure occurred in the training set was calculated.[14] This statistic was calculated for each of the three tune features (diatonic intervals, pitch, and chromatic intervals). For each individual, regression slopes using either one of these three repetition structure statistics as the only predictor were calculated. Table 7 shows the mean standardized beta values for the experimental and the control groups. From the 95% confidence intervals, it can be seen that none of these coefficients were significantly different from zero. Participants' liking responses were, therefore, not influenced by repetition structure.

Discussion

The aim of Experiment 3 was to test whether participants could learn an inversion rule that was determined solely by interval magnitude, rather than contour, because this would rule out any possible discrimination strategies based on learning local dependencies of contour patterns. The results once again showed that,

[11] There was no difference in musical experience between participants in the control group and the experimental group, $\chi^2 = 1.06$, $p = .21$.

[12] A full list of the material can be found on http://www.lifesci.sussex.ac.uk/home/Gustav_Kuhn/Kuhn_DienesJEP_LMC2006/index.htm

[13] The removal of the four tunes led to a slight imbalance in the test material.

[14] A full list of the repetition structures can be found on http://www.lifesci.sussex.ac.uk/home/Gustav_Kuhn/Kuhn_DienesJEP_LMC2006/index.htm

Table 7
Mean Standardized Beta Coefficients for the Regression Analysis Using Repetition Structure Chromatic Intervals, Repetition Structure Diatonic Intervals, and Repetition Structure Pitches as Independent Predictor Variables

Group	Variable	*M*	*SD*	LCI	UCI
Experimental	Rep. structure diatonic intervals	0.003	0.151	−0.040	0.047
Experimental	Rep. structure pitches	0.024	0.133	−0.015	0.062
Experimental	Rep. structure chromatic intervals	−0.045	0.166	−0.093	0.003
Control	Rep. structure diatonic intervals	0.003	0.153	−0.041	0.048
Control	Rep. structure pitches	−0.036	0.146	−0.078	0.006
Control	Rep. structure chromatic intervals	−0.002	0.140	−0.043	0.038

Note. Regression analysis was carried out for each participant individually, and the coefficients were averaged across participants. LCI = lower 95% confidence interval; UCI = 95% confidence interval; Rep. = repetition.

using the liking ratings, participants in the experimental group had learned to discriminate between grammatical and ungrammatical tunes. However, when the same participants were unable to make this discrimination, they were asked to give rule judgments. These results suggest that the acquired knowledge may have been below an objective threshold of awareness.

Although great care was taken to balance the test tunes in terms of sequential dependencies in the presented experiments, no special attention was paid to repetition structures. Subsequent analysis of the material revealed that for most of the test sets the repetition structures of the grammatical tunes, compared with ungrammatical tunes, were more similar to the training tunes. The grammatical tunes tended to have less repeating intervals than the ungrammatical tunes. This meant that if participants were sensitive toward the repetition structures, the sequential representations of this structure could be used to successfully discriminate between grammatical and ungrammatical items. Several studies have shown that participants in transfer studies are sensitive toward patterns of repeating elements (Brooks & Vokey, 1991; Gomez et al., 2000; Mathews & Roussel, 1997; Tunney & Altmann, 2001). Gomez et al. trained participants on a set of letter strings that contained no repetitions. This means that all items had the same repetition structure. Although participants acquired considerable knowledge about the sequential dependencies, they failed to transfer this knowledge to a novel vocabulary. These results were used to demonstrate that participants' knowledge used for transfer predominantly took the form of repetition structures. The influence of repetition structure in Experiment 3 was investigated using a regression analysis. The results from this analysis showed that participants' liking ratings were unaffected by the repetition structure, thus allowing us to exclude this explanation.

Objective threshold. In all of the test sets and experiments in which grammaticality was independent of chunks, participants in the experimental group successfully discriminated between grammatical and ungrammatical items using liking ratings, while their classification performance was at chance and no better than the control group. The knowledge demonstrated in these conditions may, therefore, have been below an objective threshold of awareness. However, strictly speaking, knowledge is only below an objective threshold if the indirect test is more sensitive than the direct test (Reingold & Merikle, 1988). To make a direct comparison between the indirect and the direct tests, classification and liking ratings were converted into z scores. For the classification data, z scores were calculated by subtracting the mean number of "yes" responses given for ungrammatical tunes from that given for grammatical tunes and dividing this difference by the mean standard deviation for that individual. A positive z score, therefore, represents above-chance discrimination performance. The same method was used to calculate z scores for the liking responses by using mean liking ratings rather than "yes" responses.

Table 8 shows the z scores for liking and discrimination responses. An ANOVA with group (experimental vs. control)[15] and experiment (abstract set vs. Experiment 2 vs. Experiment 3) as between-subjects variables and task (liking vs. classification) as within-subjects variable on z scores found no significant Group × Test × Experiment interaction, $F(2, 155) = 0.842$.$MSE = 0.077$, $p = .43$, and no significant effects of experiment, $F(2, 218) = 1.50$, $MSE = 0.105$, $p = .27$, or test $F(1, 218) = 0.269$, $MSE = 0.0247$, $p = .61$. However, there was a significant effect of group, $F(1, 218) = 8.21$, $MSE = 0.576$, $p = .005$, implying that the experimental group performed significantly better than the control group. There was no significant Group × Experiment, $F(2, 218) = 0.47$, $MSE = 0.038$, $p = .63$, or Experiment × Test, $F(2, 218) = 0.542$, $MSE = 0.380$, $p = .58$, interaction. However, there was a significant Group × Test interaction, $F(1, 218) = 8.12$, $MSE = 0.745$, $p = .005$. Furthermore, the breakdown of this interaction showed that the experimental group had significantly higher z scores on the liking task than on the classification task, $t(111) = 1.81$, $p = .033$, one-tailed, implying that these individuals had a significantly better discrimination performance on the liking task than on the classification task, thus fulfilling the objective threshold criteria. For participants in the control group, this difference was also significant, $t(111) = -2.00$, $p = .047$, but in the opposite direction. These results, therefore, show that the indirect test was more sensitive than the direct test, thus fulfilling the objective threshold criterion.

Participants' liking ratings were measured using a 9-point liking scale, and their grammaticality judgments were measured on a dichotomous yes–no scale. To increase the sensitivity of the grammaticality judgments, the yes–no responses were converted to a

[15] Participants in Experiments 1A and 1B were matched in terms of running order to allow a repeated measures analysis of task on all the data. Conversely, treating task as between subjects on all the data also gives $p < .05$, one-tailed.

Table 8
Mean z Scores Measuring Discrimination Abilities Between Grammatical and Ungrammatical Items

	Liking		Classification binary		Classification 10-point scale	
Experiment/group	*M*	*SD*	*M*	*SD*	*M*	*SD*
1a, 1b						
Experimental	0.144	0.361	0.045	0.445	0.010	0.486
Control	−0.128	0.396	0.071	0.396	0.044	0.431
2						
Experimental	0.104	0.262	0.027	0.298	0.069	0.359
Control	−0.031	0.288	0.071	0.244	0.075	0.260
3						
Experimental	0.061	0.255	0.009	0.243	0.017	0.274
Control	−0.055	0.253	−0.027	0.240	−0.012	0.226
Pooled						
Experimental	0.092	0.274	0.022	0.299	0.038	0.345
Control	−0.055	0.291	0.029	0.271	0.033	0.278

Note. The *z* scores were calculated for the liking and the classification (binary responses) and classification (10-point scale) responses. Table includes *z* scores for Experiments 1a, 1b, 2, and 3.

10-point scale using the five bins of the confidence ratings (50–60%, 61–70%, 71–80%, 81–90%, 91–100%). The "yes" responses were coded as positive (+1 to + 5) and "no" responses as negative (−1 to −5). The new values were calculated by multiplying = 1/–1 (for yes–no) with the bin order. For example, a "yes" response with a 50–59% confidence rating was coded as 1, whereby a no response with a 81–90% confidence rating was coded as −4. Using these new ratings, *z* scores for the grammaticality judgments were calculated using the same method used previously for the liking ratings.

An ANOVA with group (experimental vs. control) and experiment (abstract set vs. Experiment 2 vs. Experiment 3) as between-subjects variables and task (liking vs. classification 10 point scale) as within-subjects variable on *z* scores found no significant Group × Test × Experiment interaction, $F(2, 218) = 0.072$.$MSE = 0.701$, $p = .50$, no significant effects of experiment, $F(2, 218) = 1.73$, $MSE = 0.132$, $p = .18$, or test, $F(1, 218) = 0.272$, $MSE = 0.028$, $p = .60$. However, there was a significant effect of group, $F(1, 218) = 8.24$, $MSE = 0.630$, $p = .004$, implying that the experimental group performed significantly better than the control group. There was no significant Group × Experiment, $F(2, 218) = 0.239$, $MSE = 0.076$, $p = .79$, or Experiment × Test, $F(2, 218) = 0.016$, $MSE = 0.016$, $p = .86$, interaction. However, most importantly, there was a significant Group × Test interaction, $F(1, 218) = 6.69$, $MSE = 0.686$, $p = .01$. These results, therefore, demonstrate that the liking ratings were more sensitive regardless of the scales used.

General Discussion

The aim of the experiments presented here was to investigate whether implicit learning of musical rules could go beyond the learning of chunks of adjacent elements. Participants' awareness was evaluated both with regard to subjective and objective thresholds. The criteria of the objective threshold were based on Reingold and Merikle (1988), whereby knowledge is implicit if it influences participants' behavior on an indirect test while being inaccessible on a direct test. The indirect test was, therefore, used to assess the presence of knowledge, and the direct test was used to measure participants' awareness of this knowledge. The direct test allows for an asymmetric inference in this respect; performing above chance is not necessarily conscious knowledge, but chance performance is good evidence of a lack of conscious knowledge. However, knowledge was only claimed to be below an objective threshold of awareness if the indirect test was more sensitive than the direct test. If participants performed above chance on the direct test, confidence ratings were used to assess their metaknowledge and evaluate their awareness in respect of their subjective level of awareness. This subjective threshold of awareness was based on the theoretical framework put forward by Dienes and Perner (1999) and made use of the zero-correlation and guessing criteria.

In Experiments 1A and 1B, it was shown that participants acquired knowledge about the fragment structures in terms of either pitch or interval chunks, which was elicited using both direct and indirect tests. Because participants performed above chance on the classification task, their knowledge did not satisfy the objective threshold criterion of unconscious knowledge. Participants' level of awareness was, therefore, further analyzed with subjective measures of conscious knowledge. Although participants were asked to give verbal descriptions of the strategies used to discriminate the test tunes, these reports provided no valuable insight into their declarative knowledge. However, by using the zero-correlation criterion, it was shown that, for participants in the experimental group, there was a positive correlation between their confidence ratings and their classification performance, thus suggesting the use of conscious knowledge. These results coincide with those of Kuhn and Dienes (in press) demonstrating that learning about interval and pitch chunks led to conscious knowledge in this context. The positive zero-correlation criterion does not rule out the fact that some knowledge could be implicit. However, participants did not perform significantly different from chance when they were guessing, thus providing no evidence to suggest that they had acquired any implicit knowledge of the chunks.

The main aim of the experiments presented here was to investigate whether implicit learning could go beyond the learning of chunks of adjacent elements. Chunking models such as the competitive chunker (Boucher & Dienes, 2003; Servan-Schreiber & Anderson, 1990) or the parser (Perruchet & Vinter, 1998) become sensitive toward local dependencies and would, therefore, perform well on the material used in the exemplar and the fragment set, because grammaticality was associated with differences in chunk regularities. In the abstract set (Experiment 1A and 1B) and, in contrast, in Experiments 2 and 3, all test items were created from a set of bigrams that never occurred in the training phase. Because these bigrams have never been encountered before, chunking models prima facie would have no knowledge about these bigrams and, therefore, would be unable to apply any knowledge gained from the training items to this set of tunes. The fact that participants were able to discriminate between grammatical and ungrammatical items on these test sets implies that participants learned more than merely bigrams. One possible counterargument is that people will not perfectly perceive the different pitch or interval *n*-grams, and some new *n*-grams may be systematically confused with some old *n*-grams. Alternatively, some new *n*-grams might simply be seen as similar to some old *n*-grams. Such confusions or similarities may allow a chunking mechanism to get a handle on the material in the abstract set. The similarity of a test bigram with a training bigram could be represented by a parameter s, $0 \leq s \leq 1$, where $s = 1$ for the similarity of a bigram with itself. Performance of a chunker would deteriorate as s went to zero. Thus, a chunker would do better on a chunk set than an abstract set, other things being equal. This was not the pattern observed in Experiment 1. In sum, the results are not plausibly explained by chunking models of implicit learning.

Several studies have shown that people can incidentally learn nonlocal dependencies (Creel et al., 2004; Gomez, 2002; Newport & Aslin, 2004), the results of which would be rather challenging for chunking mechanisms. However, in these studies, the role of awareness was never directly addressed, thus making it difficult to interpret them in terms of implicit learning. In the current experiments, the fact that participants failed to distinguish between grammatical and ungrammatical items when grammaticality was not associated with differences in chunk strength indicated that the knowledge used for this discrimination may have been below an objective threshold of awareness. A direct comparison between participants' liking and their classification responses showed that the indirect test was more sensitive than the direct test. These results show that the knowledge used to distinguish between grammatical and ungrammatical items was below an objective threshold of awareness.

The results demonstrate the presence of unconscious knowledge only to the extent that the direct and indirect tests tested for the same knowledge contents (Dulany, 1962; Shanks & St. John, 1994). To illustrate the problem, consider Whittlesea and Price's (2001) argument that tests of affective judgment bias people to look at global properties of stimuli, whereas difficult tests of recognition bias people to adopt analytical strategies. The problem would arise for the existing stimuli if people were willing to base their liking on conscious knowledge of the presence of inversion but, when asked to search for rules, were biased to consider other properties of the strings. This argument can, in principle, never be ruled out, but its plausibility in this case rests on the plausibility of the claim that participants who were asked to search for musical rules and who are consciously aware of the inversion structure (or a correlated structure) in the stimuli nonetheless regard the inversion rule (or its correlate) as irrelevant to the task of searching for musical rules. The task of the critic is to make this claim plausible.

The idea that knowledge about chunks can be elicited using both direct and indirect tests, while knowledge about rules that are independent of chunks can only be elicited using affective ratings, is somewhat supported by previous studies looking at implicit learning of musical rules. In the study by Kuhn and Dienes (in press), participants' knowledge about the training items was measured using classification responses, similar to the those used in the current study. Their results showed that participants' classification responses were influenced solely by chunks rather than the inversion rule. Furthermore, Dienes and Longuet-Higgins (2004) showed that participants with a special interest in serialist music could learn musical transformations that were independent of chunks. In their discrimination task, participants were told that the melodies were genuine compositions from music students consisting of a theme–reply pairs and that for half the pairs the themes and replies had been swapped, and the other half were genuine compositions from the students, and they had to classify which was which. The task demands for this study are more similar to the task demands in the affective ratings rather than the classification task, further supporting the view that knowledge about rules that are independent of chunks can only be elicited using indirect tests with musical stimuli.

Transfer effects in which the terminal elements of the grammar are changed between the training and the test phase have been used to argue that participants acquire knowledge about the abstract structure of the grammar that is independent of the bigram structure. This interpretation would be a serious challenge to chunking models of implicit learning. However, the idea that these studies truly demonstrate implicit learning of an abstract rule has been criticized on the grounds that the abstraction of the rule may take place during the test rather than the training phase (Brooks & Vokey, 1991; Redington & Chater, 1996; Vokey & Brooks, 1992). One line of evidence supporting this view comes from studies using indirect measures. If the abstraction of the rule occurs automatically during the training phase, participants should be able to automatically apply this knowledge to the transfer set (Whittlesea & Dorken, 1997) regardless of whether knowledge is measured using an indirect or direct test. Although several studies have shown that the mere exposure effect could be generalized to a new set of grammatical letter strings (Gordon & Holyoak, 1983; Newell & Bright, 2001, 2003; Whittlesea & Dorken, 1997), this effect disappeared when the vocabulary of terminal elements was changed (Newell & Bright, 2001; Whittlesea & Dorken, 1997). The fact that in the current experiment the mere exposure effect was generalized to the test sets implies that the rules were learned automatically and during the training phase rather than the test phase.

The results from this study have shown that implicit learning could go beyond the learning of chunks of adjacent elements. If implicit learning does not merely take to form of learning about chunks, what did people learn? Several studies have shown that people in artificial grammar-learning tasks become sensitive toward the repetition structure of letter strings (Brooks & Vokey, 1991; Gomez et al., 2000; Mathews & Roussel, 1997). This type of knowledge is particularly important in transfer tasks (Brooks & Vokey, 1991; Gomez, 1997). Although no special attention was paid to the repetition structure when designing the material, subsequent regression analysis showed that participants' liking ratings were independent of the repetition structures, thus excluding this possibility. Our results suggest that participants learned some form of nonlocal dependencies. This nonlocal dependency could take two rather different forms. It is possible that individuals learned a nonlocal value–value mapping between pitches or diatonic intervals. For example, if the first note is a D, the fifth note must be a B. This mapping is similar to the rules used in the biconditional grammar-learning tasks (Johnstone & Shanks, 2001; Mathews et al., 1989; Shanks et al., 1997). However, it is also feasible that individuals learned a variable –variable mapping (i.e., in the form of operations over variables; Marcus, 2001). That is, participants learned that the inversion is formed by multiplying the diatonic intervals by -1 or subtracting each pitch number from a particular constant. However, the results presented here do not allow us to distinguish between the two possibilities. Whether people have learned a variable–variable mapping or a value–value mapping, future work could investigate whether the knowledge generalizes to different lengths of tunes; if it does, that would be evidence that people have learned not just more than chunks but more than a finite-state grammar, in fact more than just a context-free grammar (Dienes & Longuet-Higgins, 2004).

We went to great lengths here to show that implicit learning can go beyond the learning of chunks, and we illustrated how these results challenge current models of implicit learning that assume that implicit learning merely takes the form of chunks. This, of course, poses the question as to what alternative computational model can account for results presented here. One alternative model would be the two-layer feed-forward autoassociators proposed by Dienes (1992). In this model, the input of the entire sequence is represented across a set of input nodes, and the error in predicting the same activation in the output units is used as a measure of grammaticality. The same interval occurring in different positions is coded as different entities, which means that these autoassociators are insensitive to the difference between local and nonlocal dependencies. This type of autoassociator should, therefore, have no problems in learning the type of material used in these experiments. However, these models are too insensitive to the difference between local and nonlocal dependencies; people are, in fact, sensitive to the distinction (e.g., Kinder, 2000). The simple recurrent network (SRN) has become one of the most influential models of implicit learning. SRNs learn to predict the next element in a sequence. Although SRNs are particularly sensitive to transient probabilities between successive elements, one of the features that makes them particularly interesting is their ability to store information about longer dependencies (Cleeremans & McClelland, 1991; Rodriguez, 2001, 2003; Servan-Schreiber, Cleeremans, & McClelland, 1991). However, these nonlocal dependencies are learned much more slowly, and it is not clear whether the material used in current experiments could be learned. Timmermann and Cleeremans (2000) investigated whether the SRN could learn the biconditional rule used by Shanks et al. (1997). After very extensive training (3,000 epochs), the SRN was capable of distinguishing between grammatical and ungrammatical items. However, from their results, it is not clear whether the network learned the nonlocal mapping or whether the correct discrimination was based on irregularities in the test material. It, therefore, remains to be seen whether the SRN could learn the inversion rule used in the current experiments.

References

Altmann, G. T. M., Dienes, Z., & Goode, A. (1995). Modality independence of implicitly learned grammatical knowledge. *Journal of Experimental Psychology: Learning Memory & Cognition, 21,* 899–912.

Berlyne, D. E. (1970). Novelty, complexity, and hedonic value. *Perception & Psychophysics, 8,* 279–286.

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues.* Hove, UK: Erlbaum.

Bigand, E., Perruchet, P., & Boyer, M. (1998). Implicit learning of an artificial grammar of musical timbres. *Cahiers De Psychologie Cognitive-Current Psychology of Cognition, 17,* 577–600.

Bishop, C. M. (1996). *Neural networks for pattern recognition.* Oxford, UK: Clarendon Press.

Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin, 106,* 265–289.

Bornstein, R. F., & D'Agostino, P. R. (1992). Stimulus recognition and the mere exposure effect. *Journal of Personality and Social Psychology, 63,* 545–552.

Bornstein, R. F., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition, 12,* 103–128.

Boucher, L., & Dienes, Z. (2003). Two ways of learning associations. *Cognitive Science, 27,* 807–842.

Brooks, L. R. (Ed.). (1978). *Nonanalytical concept formation and memory for instances.* Hillsdale, NJ: Erlbaum.

Brooks, L. R., & Vokey, J. R. (1991). Abstract analogies and abstracted grammars: Comments on Reber (1989) and Mathews et al. (1989). *Journal of Experimental Psychology: General, 120,* 316–323.

Cheesman, J., & Merikle, P. M. (1984). Priming with and without awareness. *Perception and Psychophysics, 36,* 387–395.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences, 2,* 406–416.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General, 120,* 235–253.

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning Memory, and Cognition, 30,* 1119–1130.

Cross, I. (1997). Pitch schemata. In I. S. Deliege (Ed.), *Perception and cognition of music* (pp. 357–390). Hove, UK: Psychology Press.

Dienes, Z. (1992). Connectionist and memory-array models of artificial grammar learning. *Cognitive Science, 16,* 41–79.

Dienes, Z. (2004). Assumptions of subjective measures of unconscious mental states: Higher order thoughts and bias. *Journal of Consciousness Studies, 11,* 25–45.

Dienes, Z., & Altmann, G. T. (Eds.). (1997). *Transfer of implicit knowledge across domains. How implicit and how abstract?* Oxford, UK: Oxford University Press.

Dienes, Z., & Altmann, G. T. (2003). Measuring learning using an untrained control group: Comments on R. Reber and P. Perruchet. *Quarterly Journal of Experimental Psychology: Series A. Human Experimental Psychology, 56,* 117–123.

Dienes, Z., Altmann, G. T. M., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning Memory, and Cognition, 21,* 1322–1338.

Dienes, Z., Broadbent, D., & Berry, D. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 875–887.

Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learnt? *Cognitive Science, 28,* 531–558.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioural and Brain Sciences, 22,* 735–755.

Dienes, Z., & Perner, J. (2001). When knowledge is unconscious because of conscious knowledge and vice versa. Proceedings of the 23rd Annual Conference of the Cognitive Science Society (pp. 255–260), August 1–4, Edinburgh, Scotland. Mahwah, NJ: Erlbaum.

Dienes, Z., & Perner, J. (Eds.). (2004). *Assumptions of a subjective measure of consciousness: Three mappings.* Amsterdam: John Benjamins.

Dienes, Z., & Scott, R. (2005). Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychological Research, 69,* 338–351.

Dowling, W. J. (1971). Recognition of inversions of melodies and melodic contours. *Perception & Psychophysics, 9,* 348–349.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* New York: Academic Press.

Dulany, D. E. (Ed.). (1962). *The place of hypotheses and intentions: An analysis of verbal control in verbal conditioning.* Durham, NC: Duke University Press.

Dulany, D. E., Carlson, R. A., & Dewey, G. I. (1984). A case of syntactical learning and judgment: How conscious and how abstract. *Journal of Experimental Psychology: General, 113,* 541–555.

Frances, R. (1988). *The perception of music* (W. J. Dowling, Trans.). Hillsdale, NJ: Erlbaum.

Gomez, R. L. (1997). Transfer and complexity in artificial grammar learning. *Cognitive Psychology, 33,* 154–207.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13,* 431–436.

Gomez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition, 28,* 253–263.

Gordon, P. C., & Holyoak, K. J. (1983). Implicit learning and generalization of the mere exposure effect. *Journal of Personality and Social Psychology, 45,* 492–500.

Johnstone, T., & Shanks, D. R. (1999). Two mechanisms in implicit artificial grammar learning? Comment on Meulemans and Van der Linden (1997). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 524–531.

Johnstone, T., & Shanks, D. R. (2001). Abstractionist and processing accounts of implicit learning. *Cognitive Psychology, 142,* 61–112.

Kelley, C. M., Burton, M. A., Kato, T., & Akamatsu, S. (2001). Incidental learning of real-world regularities. *Psychological Science, 12,* 86–89.

Kinder, A. (2000). The knowledge acquired during artificial grammar learning: Testing the predictions of two connectionist models. *Psychological Research, 63,* 95–105.

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 79–91.

Knowlton, B. J., & Squire, L. R. (1996). Artificial grammar learning depends on implicit acquisition of both abstract and exemplar-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 169–181.

Krumhansl, C. L. (1991). Music psychology: Tonal structures in perception and memory. *Annual Review of Psychology, 42,* 277–303.

Krumhansl, C. L., Toivanen, P., Eerola, T., Toiviainen, P., Jarvinen, T., & Louhivuori, J. (2000). Cross-cultural music cognition: Cognitive methodology applied to North Sami yolks. *Cognition, 76,* 13–58.

Kuhn, G., & Dienes, Z. (in press). Differences in the types of musical regularities learnt in incidental and intentional learning conditions. *Quarterly Journal of Experimental Psychology.*

Kunst-Wilson, W. R., & Zajonc, R. B. (1980). Affective discrimination of stimuli that cannot be recognised. *Science, 207,* 557–558.

Lee, A. Y. (2001). The mere exposure effect: An uncertainty reduction explanation revisited. *Personality and Social Psychology Bulletin, 27,* 1255–1266.

Lewicki, P. (1986). *Nonconscious social information processing.* New York: Academic Press.

Longuet-Higgins, H. C. (1987). *Mental processes: Studies in cognitive science.* Cambridge, MA: MIT Press.

Manza, L., & Bornstein, R. F. (1995). Affective discrimination and the implicit learning process. *Consciousness and Cognition, 4,* 399–409.

Manza, L., & Reber, A. S. (1997). Representing artificial grammars: Transfer across stimulus forms and modalities. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 73–106). Oxford, UK: Oxford University Press.

Marcus, G. (2001). *The algebraic mind.* Cambridge, MA: MIT Press.

Mathews, R. C., Buss, R. R., Stanley, W. B., Blanchardfields, F., Cho, J. R., & Druhan, B. (1989). Role of implicit and explicit processes in learning from examples: A synergistic effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 1083–1100.

Mathews, R. C., & Roussel, L. G. (1997). Abstractness of implicit knowledge: A cognitive evolutionary perspective. In D. Berry (Ed.), *How implicit is implicit learning?* (pp. 13–47). New York: Oxford University Press.

Meulemans, T., & Van der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1007–1028.

Meyer, M. (1903). Experimental studies in the psychology of music. *American Journal of Psychology, 14,* 456–478.

Newell, B. R., & Bright, J. E. H. (2001). The relationship between the structural mere exposure effect and the implicit learning process. *Quarterly Journal of Experimental Psychology: Series A. Human Experimental Psychology, 54,* 1087–1104.

Newell, B. R., & Bright, J. E. H. (2002). Well past midnight: Calling time on implicit invariant learning? *European Journal of Cognitive Psychology, 14,* 185–205.

Newell, B. R., & Bright, J. E. H. (2003). The subliminal mere exposure effect does not generalize to structurally related stimuli. *Canadian Journal of Experimental Psychology, 57,* 61–68.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: 1. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

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

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge. *Journal of Experimental Psychology: General, 119,* 264–275.

Perruchet, P., & Vinter, A. (1998). PARSER: A model of word segmentation. *Journal of Memory and Language, 39,* 246–263.

Reber, A. S. (1967). Implicit learning of artificial grammar. *Journal of Verbal Learning & Verbal Behaviour, 6,* 855–863.

Reber, A. S. (1989). Implicit learning and tactic knowledge. *Journal of Experimental Psychology: General, 118,* 219–1989.

Redington, M., & Chater, N. (1996). Transfer in artificial grammar learning: A reevaluation. *Journal of Experimental Psychology: General, 125,* 123–138.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception & Psychophysics, 44,* 563–575.

Reingold, E. M., & Merikle, P. M. (1993). Theory and measurements in the study of unconscious processes. In M. Davies & G. W. Humphreys (Eds.), *Consciousness.* Cambridge, MA: Blackwell.

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation, 13,* 2093–2118.

Rodriguez, P. (2003). Comparing simple recurrent networks and n-grams in a large corpus. *Applied Intelligence, 19,* 39–50.

Rosenthal, D. M. (Ed.). (2002). *Consciousness and higher-order thought.* New York: Macmillan.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8,* 101–105.

Schoenberg, A. (1941). *Style and idea: Selected writings of Arnold Schoenberg.* London: Faber and Faber.

Seamon, J. G., Brody, N., & Kauff, D. M. (1983). Affective discrimination of stimuli that are not recognized. Effect of shadowing, masking, and cerebral laterality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 544–555.

Seamon, J. G., McKenna, P. A., & Binder, N. (1998). The mere exposure effect is differentially sensitive to different judgment tasks. *Consciousness and Cognition, 7,* 85–102.

Servan-Schreiber, D., Cleeremans, A., & McClelland, J. L. (1991). Graded state machines: The representation of temporal contingencies in simple recurrent networks. *Machine Learning, 7,* 161–193.

Servan-Schreiber, E., & Anderson, J. R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 592–608.

Shanks, D. R., Johnstone, T., & Staggs, L. (1997). Abstraction processes in artificial grammar learning. *Quarterly Journal of Experimental Psychology: Section A. Human Experimental Psychology, 50,* 216–252.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning-systems. *Behavioural and Brain Sciences, 17,* 367–395.

Shepard, R. N. (1982). Structural representation of musical pitch. In D. Deutsch (Ed.), *The psychology of music.* London: Academic Press.

Smith, D., & Witt, J. (1989). Spun steel and stardust: The rejection of contemporary compositions. *Music Perception, 7,* 169–186.

Tillmann, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107,* 885–913.

Timmermann, B., & Cleeremans, A. (2000). Rules vs. statistics in biconditional grammar learning: A simulation based on Shanks et al. (1997). *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society (pp. 947–952).* Mahwah, NJ: Erlbaum.

Tunney, R. J., & Altmann, G. T. M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 614–639.

Tunney, R. J., & Shanks, D. R. (2003). Subjective measures of awareness and implicit cognition. *Memory & Cognition, 31,* 1060–1071.

Verveer, E., Barry, H., & Bousfield, W. (1933). Change in affectivity with repetition. *American Journal of Psychology, 45,* 130–134.

Vokey, J. R., & Brooks, L. R. (1992). Salience of item knowledge in learning artificial grammars. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 328–344.

Whittlesea, B. W. A., & Dorken, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing account of implicit learning. *Journal of Experimental Psychology: General, 122,* 227–248.

Whittlesea, B. W. A., & Dorken, M. D. (1997). Implicit learning: Indirect, not unconscious. *Psychonomic Bulletin & Review, 4,* 63–67.

Whittlesea, B. W. A., & Price, J. R. (2001). Implicit/explicit memory versus analytic/nonanalytic processing: Rethinking the mere exposure effect. *Memory & Cognition, 29,* 234–246.

Wilson, W. (1979). Feeling more than we can know: Exposure effects without learning. *Journal of Personality and Social Psychology, 37,* 811–821.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monographs, 9,* 1–27.

Received June 29, 2004
Revision received June 16, 2005
Accepted June 22, 2005 ■