Modality-Constrained Statistical Learning of Tactile, Visual, and Auditory Sequences

Christopher M. Conway and Morten H. Christiansen
Cornell University

The authors investigated the extent to which touch, vision, and audition mediate the processing of statistical regularities within sequential input. Few researchers have conducted rigorous comparisons across sensory modalities; in particular, the sense of touch has been virtually ignored. The current data reveal not only commonalities but also modality constraints affecting statistical learning across the senses. To be specific, the authors found that the auditory modality displayed a quantitative learning advantage compared with vision and touch. In addition, they discovered qualitative learning biases among the senses: Primarily, audition afforded better learning for the final part of input sequences. These findings are discussed in terms of whether statistical learning is likely to consist of a single, unitary mechanism or multiple, modality-constrained ones.

The world is temporally bounded: Events do not occur all at once but rather are distributed in time. Therefore, it is crucial for organisms to be able to encode and represent temporal order information. One potential method for encoding temporal order is to learn the statistical relationships of elements within sequential input. This process appears to be important in a diverse set of learning situations, including speech segmentation (Saffran, Newport, & Aslin, 1996), learning orthographic regularities of written words (Pacton, Perruchet, Fayol, & Cleeremans, 2001), visual processing (Fiser & Aslin, 2002), visuomotor learning (e.g., serial reaction time tasks; Cleeremans, 1993) and nonlinguistic, auditory processing (Saffran, Johnson, Aslin, & Newport, 1999). Not only human adults but also infants (Gomez & Gerken, 1999; Kirkham, Slemmer, & Johnson, 2002; Saffran, Aslin, & Newport, 1996) and nonhuman primates (Hauser, Newport, & Aslin, 2001) are capable of statistical learning.

Noting such widespread examples of statistical learning, many researchers—either implicitly or explicitly—view statistical learning as a single, domain-general phenomenon (e.g., Kirkham et al., 2002). Although it may be true that statistical learning across different domains is based on similar computational principles, it is also likely that modality constraints exist that may differentially affect such processing. For instance, traditionally, vision and audition have been viewed as spatial and temporal senses, respec-

tively (Kubovy, 1988). Empirical evidence from perceptual and temporal processing experiments supports such a distinction between vision and audition (e.g., Glenberg & Swanson, 1986; Mahar, Mackenzie, & McNicol, 1994). However, it is currently unknown whether and how these modality constraints affect the learning of statistical relationships between elements contained within sequential input.

This article explores potential modality constraints affecting statistical learning. Experiment 1 investigates statistical learning in three sensory modalities: touch, vision, and audition. Experiment 1A provides the first direct evidence that touch can mediate statistical learning. Experiments 1B and 1C compare learning in two additional sensory modalities, vision and audition. Although commonalities exist, we find initial evidence for a striking difference in auditory statistical learning compared with tactile and visual learning. We follow up with Experiment 2, designed to control perceptual and training effects as well as to tease apart potential learning sensitivities uncovered in the first experiment. The results of Experiment 2 provide further evidence that modality constraints affect statistical learning. We discuss these results in relation to basic issues of cognitive and neural organization— namely, to what extent statistical learning might consist of a single or multiple neural mechanisms.

Statistical Learning of Sequential Input

Statistical learning appears to be a crucial learning ability. For instance, making sense of visual scenes may require the extraction of statistical components (e.g., Fiser & Aslin, 2001). Another domain in which statistical learning likely plays an important role is the encoding of sequential input (Conway & Christiansen, 2001). Artificial grammar learning (AGL; Reber, 1967) is a paradigm widely used for studying such statistical learning.[1] AGL experiments typically use finite-state grammars to generate the

---

[1] The serial reaction time (SRT) task is another common method for exploring the learning of sequential regularities. The SRT paradigm differs from AGL in that the behavioral measure for the former is reaction time, whereas that for the latter is classification accuracy.

stimuli. In such grammars, a transition from one state to the next produces an element of the sequence. For example, by passing through the nodes S1, S2, S2, S4, S3, S5 of Figure 1, one generates the "legal" sequence 4–1–3–5–2.

In the AGL paradigm, participants observe a subset of legal training sequences (i.e., sequences that are generated from the artificial grammar), after which the participants typically display learning of sequential structure as evidenced by their ability to classify novel sequences as being legal or illegal. Additionally, they often have difficulties verbalizing the distinction between legal and illegal stimuli, a finding that originally prompted Reber (1967) to describe the learning as implicit.

The nature of the cognitive processes underlying AGL has been the subject of much debate, leading to the proposal of several different theories. The abstractive view sees AGL as a process that encodes and extracts the abstract rules of the grammar (e.g., Reber, 1993). Two alternative accounts stand in contrast to the abstractive view, proposing that instead of abstract knowledge, participants learn particular features of the training items. The exemplar-based view posits that the stimuli themselves are encoded and stored in memory (e.g., Vokey & Brooks, 1992): When participants make classification judgments at test, they compare the test sequences with their memory of the stored exemplars and make their decision on the basis of similarity. The fragment-based view posits that participants learn small fragments or chunks of information, consisting of pairs (*bigrams*) and triples (*trigrams*) of elements (e.g., Perruchet & Pacteau, 1990). Participants use these chunks of information to help them classify novel input.

Although there has been disagreement as to which theory is correct, there is considerable evidence suggesting that the learning of fragment information is a crucial aspect of AGL[2] (e.g., Johnstone & Shanks, 1999; Knowlton & Squire, 1994, 1996; Meulemans & Van der Linden, 1997; Perruchet & Pacteau, 1990; Pothos & Bailey, 2000; Redington & Chater, 1996). These experiments have shown that participants become sensitive to the fragment information contained within the training input, as quantified by specific fragment measures, which allows participants to classify novel sequences in terms of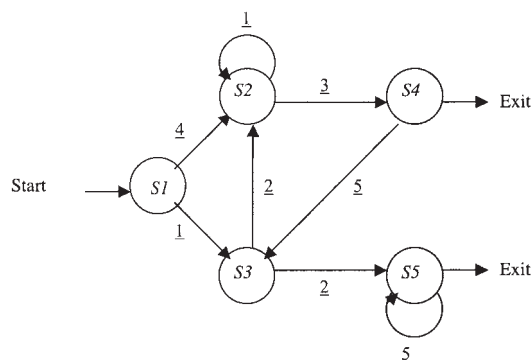 whether they conform to the same statistical regularities as the training items. Such statistical sensitivity appears to be vital for AGL tasks.

The standard AGL paradigm has been used extensively to assess visual as well as auditory (e.g., Saffran, 2000) learning. However, two issues remain relatively unexplored: Can statistical learning occur in other modalities, such as touch? And what differences in statistical learning, if any, exist among different sensory modalities? Whereas previous research generally has focused on the similarities among statistical learning in different domains (Fiser & Aslin, 2002; Kirkham et al., 2002), there are reasons to suppose that modality constraints may affect learning across the various senses. Next, we summarize evidence for such modality constraints.

Modality Constraints

Ample research testifies to the existence of modality constraints that affect the manner in which people perceive, learn, and represent information (for relevant reviews, see Freides, 1974; Penney, 1989). In this section we summarize research in the realms of serial recall, temporal acuity, and the learning of temporal and statistical patterns.

One of the most well-known modality effects—often referred to as *the* modality effect—is found in serial recall. Numerous studies attest to differences in the serial position learning curves for aurally versus visually presented verbal input (e.g., lists of spoken or written words). Specifically, there appears to be a stronger recency effect (i.e., better recall of final elements in a list) for auditory as compared with visual material (Crowder, 1986; Engle & Mobley, 1976). A number of theories have attempted to explain this modality effect, such as the traditional account supposing that a precategorical acoustic storage exists for auditory material (Crowder & Morton, 1969) or that the auditory modality benefits from better temporal coding (e.g., Glenberg & Fernandez, 1988). Beaman (2002) showed that under certain conditions, a stronger primacy effect (i.e., better recall of beginning elements in a list) occurs for visual as compared with auditory material. Traditional theories do not adequately explain why this might occur. Additionally, studies with nonhuman primates have shown that monkeys have opposite serial position curves for auditory and visual material (Wright, 2002), as a function of the amount of time occurring between the last element in the list and the recall test. That is, when the recall test occurs relatively soon after the list presentation, there is an auditory primacy effect and a visual recency effect; when the recall test occurs relatively late after the presentation, there is a visual primacy and an auditory recency effect. These new data suggest that different mechanisms may underlie auditory and visual serial recall, leading to qualitatively different serial position curves.

Modality differences are also apparent in low-level temporal processing tasks (e.g., Gescheider, 1966, 1967; Lechelt, 1975; Oatley, Robertson, & Scanlan, 1969; Sherrick & Cholewiak, 1986). For example, Sherrick and Cholewiak (1986) reviewed data relating to temporal acuity in touch, vision, and audition. In mea-

Figure 1. Artificial grammar adapted from Gomez and Gerken (1999), also used in the current Experiment 1. We generated legal sequences by following the paths starting at S1 and continuing until we reached an exit path. Each path generates a number (1, 2, 3, 4, or 5) that corresponds to a particular stimulus element. S = state, so that S1 and S2 refer to State 1 and State 2, and so on.

---

[2] It also appears to be the case that learners rely on other cues, such as overall similarity of test items to training exemplars, in addition to fragment information (e.g., see Pothos & Bailey, 2000).

sures of simultaneity—the ability to correctly perceive two closely occurring events—the senses have differing temporal sensitivity, with vision being the least and audition the most sensitive. Similarly, Lechelt (1975) assessed each modality in terms of numerosity, or the ability to count rapidly presented stimuli. Stimuli consisting of flashes of light, aural clicks, or finger taps were delivered for short durations (2 ms or less), with sequences of varying length (between two and nine pulses) and varying rates (between three and eight signals per second). In terms of assessing the number of signals in the sequences, participants performed best when the signals were presented aurally and worst when they were presented visually.

Likewise, studies of temporal pattern and rhythm discrimination also reveal modality differences (e.g., Collier & Logan, 2000; Garner & Gottwald, 1968; Glenberg & Jona, 1991; Handel & Buffardi, 1969; Manning, Pasquali, & Smith, 1975; Rubinstein & Gruenberg, 1971). When presented with rhythmic patterns of flashing lights or auditory stimuli, participants were much better at discriminating auditory as opposed to visual patterns (Rubinstein & Gruenberg, 1971). Learners were also better at identifying repeating sequences of binary elements (e.g., 1122121211221212) when the elements were auditory stimuli rather than visual or tactual ones (Handel & Buffardi, 1969).

There have also been hints that similar modality constraints affect AGL. Several studies have noted that performance in AGL tasks differs depending on the modality and the manner of presentation (i.e., whether material is presented simultaneously or sequentially). For instance, Gomez (1997) remarked that visual AGL proceeds better when the stimuli are presented simultaneously rather than sequentially, perhaps because a simultaneous format permits better chunking of the stimulus elements. Saffran (2002) used an AGL task to test participants' ability to learn predictive dependencies. She found that participants learned these predictive relationships best with an auditory–sequential or visual–simultaneous presentation and did poorly in a visual–sequential condition.

The evidence reviewed suggests that modality differences are present across the cognitive spectrum. These modality constraints take two main forms. First, it appears that vision and audition differ in respect to their sensitivities to the initial or final parts of sequential input. Vision may be more sensitive to initial items in a list (Beaman, 2002), whereas audition appears more sensitive to final list items (Crowder, 1986). Second, the auditory modality appears to have an advantage in the processing of sequential input, including low-level temporal processing tasks (Sherrick & Cholewiak, 1986) and pattern or rhythm discrimination (e.g., Manning et al., 1975). In a comprehensive review of the effect of modality on cognitive processing, Freides (1974) concluded that for complex tasks, audition is best suited for temporal processing, whereas vision excels at spatial tasks (for similar views, see also Kubovy, 1988; Mahar et al., 1994; Penney, 1989; Saffran, 2002). That is, audition is best at processing sequential, temporally distributed input, whereas vision excels at spatially distributed input. The touch modality appears to be adept at processing both sequential and spatial input, but not at the same level of proficiency as either audition or vision (Mahar et al., 1994).

In this article we explore in what manner these modality constraints might affect statistical learning. In the experiments, our strategy is to incorporate comparable input in three sensory con-

ditions: touch, vision, and audition. Previous researchers have claimed that statistical learning in audition and vision is the same, yet rarely has much effort been made to control experimental procedures and materials across the senses. Thus, the present experiments provide a better comparison of learning across these three modalities. We begin by investigating statistical learning in the tactile domain, a realm that has been previously ignored in AGL experiments.

Experiment 1A: Tactile Statistical Learning

The touch sense has been studied extensively in terms of its perceptual and psychophysical attributes (see Craig & Rollman, 1999), yet it has not been fully explored in relation to statistical learning. In Experiment 1A, we presented to participants tactile sequences conforming to an artificial grammar and then tested their ability to classify novel sequences. As reviewed above, studies of sequential pattern perception suggest that the touch sense ought to be capable of extracting sequential regularities in an AGL setting (e.g., Handel & Buffardi, 1969; Manning et al., 1975). This experiment attempted to verify this hypothesis.

Method

Participants

Twenty undergraduates (10 in each condition) from introductory psychology classes at Southern Illinois University participated in the experiment. Subjects earned course credit for their participation. The data from an additional 5 participants were excluded for the following reasons: prior participation in AGL tasks in our laboratory ($n = 4$), and failure to adequately follow the instructions ($n = 1$).

Apparatus

The experiment was conducted with the PsyScope presentation software (Cohen, MacWhinney, Flatt, & Provost, 1993) run on an Apple G3 PowerPC computer. Participants made their responses using an input/output button box (New Micros, Inc., Dallas, TX). Five small motors (18 mm $\times$ 5 mm), normally used in hand-held paging devices, generated the vibrotactile pulses (rated at 150 Hz). The vibration pulses were suprathreshold stimuli and easily perceived by all participants. The motors were controlled by output signals originating from the New Micros button box. These control signals were in turn determined by the PsyScope program, which allowed precise control over the timing and duration of each vibration stimulus. Figure 2 shows the general experimental setup.

Materials

The stimuli used for Experiment 1 were taken from Gomez and Gerken's (1999) Experiment 2. This grammar (see Figure 1) can generate up to 23 sequences between three and six elements in length. The grammar generates sequences of numbers. Each number from the grammar was mapped onto a particular finger (1 was the thumb, and 5 was the pinky finger). Each sequence generated from the grammar thus represents a series of vibration pulses delivered to the fingers, one finger at a time. Each finger pulse duration was 250 ms, and the pulses within a sequence were separated by 250 ms. As an illustration, the sequence 1–2–5–5 corresponds to a 250-ms pulse delivered to the thumb, a 250-ms pause, a 250-ms pulse delivered to the second finger, a 250-ms pause, a 250-ms pulse delivered to the fifth finger, a 250-ms pause, and then a final 250-ms pulse delivered to the fifth finger. Figure 3 graphically represents this sequence.
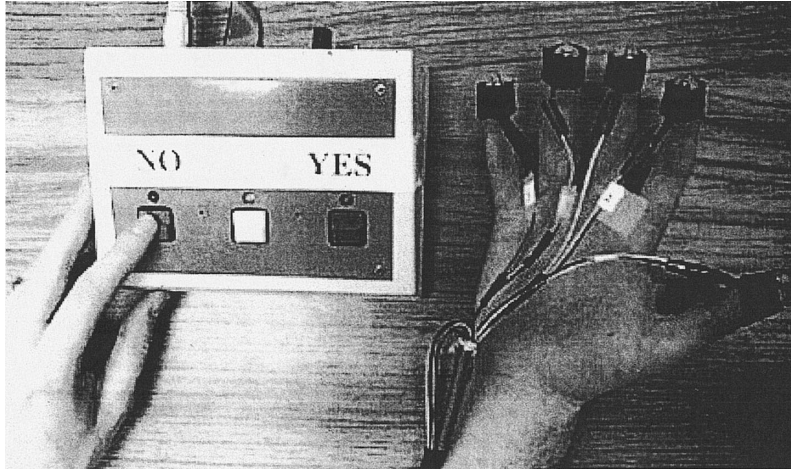
Figure 2. Vibration devices attached to a participant's hand with the button box to the side (Experiment 1A).

A total of 12 legal sequences were used for training.[3] Each of the legal sequences was used twice to formulate a set of 12 training pairs. Six pairs consisted of the same training sequence presented twice (matched pairs), whereas the remaining 6 pairs consisted of 2 sequences that differed slightly from one another (mismatched pairs). These matched and mismatched training pairs were used in conjunction with a same–different judgment task, described in detail below. The 12 training pairs are listed in Appendix A.

The test set consisted of 10 novel legal and 10 illegal sequences. Legal sequences were produced from the finite-state grammar in the normal fashion. Illegal sequences did not conform to the regularities of the grammar. The illegal sequences each began with a legal element (i.e., 1 or 4), followed by one or more illegal transitions and ending with a legal element (i.e., 2, 3, or 5). For example, the illegal sequence 4–2–1–5–3 begins and ends with legal elements (4 and 3, respectively) but contains several illegal interior transitions (4–2, 1–5, and 5–3, combinations of elements that the grammar does not allow). Therefore, the legal and illegal sequences can be described as differing from one another in terms of the statistical relationships between adjacent elements. That is, a statistical learning mechanism able to encode the possible element combinations occurring in the training set could discern which novel test sequences are illegal. For instance, by

realizing that the elements 4 and 2 never occur together in the training set, a learner could potentially discern that the novel test sequence 4–2–1–5–3 is illegal.[4] Finally, the legal and illegal test sequences were closely matched in terms of element frequencies and sequence lengths (Gomez & Gerken, 1999). All test sequences are listed in Table 1.

Procedure

Participants were assigned randomly to either a control group or an experimental group. The experimental group participated in both a training and a test phase, whereas the control group only participated in the test phase. Before beginning the experiment, all participants were assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) to determine their preferred hand. The experimenter then placed a vibration device onto each of the five fingers of the participant's preferred hand. At the beginning of the training phase, the experimental group participants were instructed that they were participating in a sensory experiment in which they would feel pairs of vibration sequences. For each pair of sequences, they had to decide whether the two sequences were the same and indicate their decision by pressing a button marked *YES* or *NO*. This match–mismatch paradigm used the 12 training pairs described earlier, listed in Appendix A. It was our intention that this paradigm would encourage participants to pay attention to the stimuli while not directly tipping them off to the nature of the statistically governed sequences.

Each pair was presented six times in random order for a total of 72 exposures. As mentioned earlier, all vibration pulses had a duration of 250 ms and were separated by 250 ms within a sequence. A 2-s pause occurred between the two sequences of each pair and after the last sequence of the pair. A prompt was displayed on the computer monitor asking for the participant's response, and it stayed on the screen until a button press was made. After another 2-s pause, the next training pair was presented. The entire training phase lasted roughly 10 min for each participant.

A recording of white noise was played during training to mask the sounds of the vibrators. In addition, the participants' hands were occluded so that they could not visually observe their fingers. These precautions
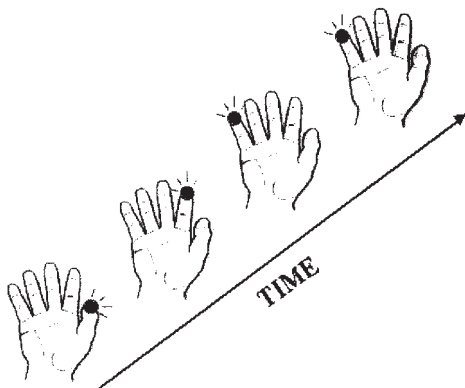
Figure 3. Graphical representation of the tactile sequence 1–2–5–5 in Experiment 1A. Each hand represents a single slice in time, whereas each black circle represents the occurrence of a vibrotactile pulse to a particular finger.

---

[3] Note that what we refer to as the training phase contained neither performance feedback nor reinforcement of any kind. Exposure phase might be a more accurate description of this part of the experiment.

[4] Note that we remain neutral as to whether such performance might occur in the presence or absence of awareness.

Table 1
Fragment Measures for Experiment 1 Test Sequences

Item	Chunk	Novel	NFP	Sim.	I-anchor	F-anchor
		Legal test sequences				
4–1–3–5–2–3	4.11	0	0	2	2.5	2.5
1–2–1–3–5–2	4.11	0	0	2	4.5	2.0
4–3–5–2–5–5	3.67	0	2	4	2.0	1.5
4–1–3–5–2–5	4.00	0	0	2	2.5	2.0
4–1–1–1–3	2.57	0	3	3	2.0	2.0
1–2–1–1–3	3.14	0	2	2	4.5	2.0
1–2–3–5–2	5.00	0	0	1	5.0	2.0
4–1–1–3	2.80	0	0	2	2.0	2.0
4–3–5–2	4.40	0	0	1	2.0	2.0
1–2–5	4.33	0	0	1	4.5	1.0
Average	3.81	0.0	0.70	2.00	3.15	1.9
		Illegal test sequences				
1–4–5–1–3–3	0.56	8	9	3	0	0
4–5–1–2–1–3	1.89	4	8	3	0	2.0
4–2–1–3–1–5	0.89	6	6	3	0	0
1–5–3–3–2–2	0.00	9	9	3	0	0
1–5–3–4–2	0.00	7	7	3	0	0
4–2–1–5–3	0.29	6	6	3	0	0
1–5–3–1–2	1.00	6	7	3	0	0
4–5–1–3	1.00	4	4	2	0	1.5
4–5–2–2	1.20	4	5	3	0	0
1–4–2	0.00	3	3	2	0	0
Average	0.68	5.70	6.40	2.80	0	0.35

Note. NFP = Novel fragment position; Sim. = similarity; I-anchor = initial anchor strength; F-anchor = final anchor strength.

were taken to ensure that tactile information alone, without help from auditory or visual senses, contributed to task performance.

Before the beginning of the test phase, the experimental group participants were told that the vibration sequences they had just felt had been generated by a computer program that determined the order of the pulses by using a complex set of rules. They were told that they would now be presented with new vibration sequences. Some of these would be generated by the same program, whereas others would not be. It was the participant's task to classify each new sequence accordingly (i.e., whether or not the sequence was generated by the same rules) by pressing a button marked either *YES* or *NO*. The control participants, who did not participate in the training phase, received an identical test task.

The 20 test sequences were presented one at a time, in random order, to each participant. The timing of the test sequences was the same as that used during the training phase (250-ms pulse duration, 250-ms interstimulus interval, and 2-s pauses before and after each sequence). The white noise recording and occluding procedures also were continued in the test phase.

At the completion of the experiment, participants were asked how they decided whether test sequences were legal or illegal. Some researchers have used such verbal reports as a preliminary indication as to whether learning proceeded implicitly or explicitly (Seger, 1994).

Results and Discussion

We assessed the training performance for the experimental participants by calculating the mean percentage of correctly classified pairs. Participants, on average, made correct match–mismatch decisions for 74% of the training trials.

However, for our purposes, the test results are of greater interest because here the participants must generalize from training expe-

rience to previously unobserved test sequences. The control group correctly classified 45% of the test sequences, whereas the experimental group correctly classified 62% of the test sequences. Following Redington and Chater's (1996) suggestions, we conducted two analyses on the test data. The first was a one-way analysis of variance (ANOVA; experimental vs. control group) to determine whether any differences existed between the two groups. The second compared performances for each group with hypothetical chance performance (50%) using single group t tests.

The ANOVA revealed that the main effect of group was significant, $F(1, 18) = 3.16$, $p < .01$, indicating that the experimental group performed significantly better than the control group. Single group t tests confirmed the ANOVA's finding. The control group's performance was not significantly different from chance, $t(9) = -1.43$, $p = .186$, whereas the experimental group's performance was significantly above chance, $t(9) = 2.97$, $p < .05$.

Finally, the participants' verbal reports suggest that they had very little explicit knowledge concerning sequence legality. Most of the experimental group participants reported basing their responses merely on whether a sequence felt familiar or similar. Several of the participants reported that they made their judgments on the basis of a simple rule (e.g., "If a sequence was four elements long, I said 'no'"). However, in each of these cases, following the rule would actually lead to incorrect judgments. None of the participants was able to report anything specific that could actually help him or her make a decision (e.g., "Certain finger combinations were not allowed, such as the fourth finger followed by the second"). On the basis of these verbal reports, we do not see evidence that the experimental group participants were explicitly aware of the distinction between legal and illegal sequences.[5]

The results show that the experimental group significantly outperformed the control group. This suggests that the experimental participants learned aspects of the statistical structure of the training sequences—in the form of adjacent element co-occurrence statistics—that allowed them to classify novel test sequences appropriately. Additionally, the participants had difficulty verbalizing the nature of sequence legality. This is the first empirical evidence of an apparently implicit, tactile statistical learning capability.

Experiments 1B and 1C: Visual and Auditory Statistical Learning

Experiment 1A showed that statistical learning can occur in the tactile domain. To compare tactile with visual and auditory learning, we conducted two additional studies. Experiments 1B and 1C assessed statistical learning in the visual and auditory domains, respectively, using the same general procedure and statistically governed stimulus set as used in Experiment 1A. For Experiment 1B, the sequences consisted of visual stimuli occurring at different spatial locations. For Experiment 1C, sequences of tones were used. Like the vibrotactile sequences, the visual and auditory stimuli were nonlinguistic, and thus participants could not rely on a verbal encoding strategy.

---

[5] We note, however, that verbal reports are not necessarily the most sensitive measure of explicit awareness, so it is still possible that explicit awareness contributed to task performance.

Figure 5. Graphical representation of the auditory sequence 1–2–5–5 in Experiment 1C.

Method

Participants

Experiment 1B. Twenty undergraduates (10 in each condition) were recruited from introductory psychology classes at Cornell University. Subjects received extra credit for their participation. The data from 3 additional participants were excluded because the participants did not adequately follow the instructions (*n* = 2) and because of equipment malfunction (*n* = 1).

Experiment 1C. An additional 20 undergraduates (10 in each condition) were recruited from introductory psychology classes at Cornell University.

Apparatus

The apparatus was the same as in Experiment 1A, except for the exclusion of the vibration devices. The auditory stimuli were generated by the SoundEdit 16 (Version 2) software for the Macintosh.

Materials

The training and test materials were identical to those of Experiment 1A (see Appendix A and Table 1). The difference was that the sequence elements were mapped onto visual or auditory stimuli instead of vibrotactile pulses. For Experiment 1B, the stimuli consisted of black squares displayed on the computer monitor in different locations (the element 1 represents the leftmost location, and 5 the rightmost). Each black square (2.6 × 2.6 cm) was positioned in a horizontal row across the middle of the screen at approximately eye level, with 2.5 cm separating each position. Participants were seated at a viewing distance of approximately 45 cm to 60 cm from the monitor.

A visual stimulus thus consisted of a spatiotemporal sequence of black squares appearing at various locations. As in Experiment 1A, each element appeared for 250 ms, and each was separated by 250 ms. Figure 4 shows a representation of the sequence 1–2–5–5.

For Experiment 1C, the stimuli consisted of pure tones of various frequencies (1 = 261.6 Hz, 2 = 277.2 Hz, 3 = 349.2 Hz, 4 = 370 Hz, and 5 = 493.9 Hz) corresponding to musical notes C, C#, F, F#, and B, respectively.[6] As in Experiments 1A and 1B, each element (tone) lasted 250 ms, and each was separated by 250 ms. Figure 5 graphically represents the sequence 1–2–5–5.
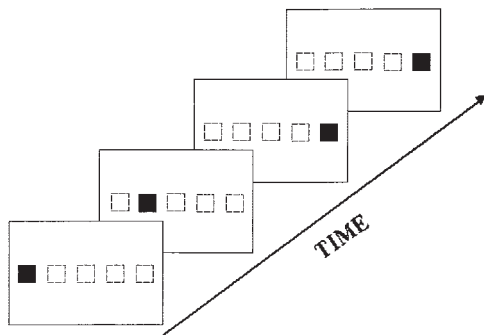

Figure 4. Graphical representation of the visual sequence 1–2–5–5 in Experiment 1B. Each of the four large rectangles represents the monitor display at a single slice in time. Filled squares represent the occurrence of a visual stimulus. Note that the dashed squares, representing the five possible stimulus element locations, were not visible to the participants.
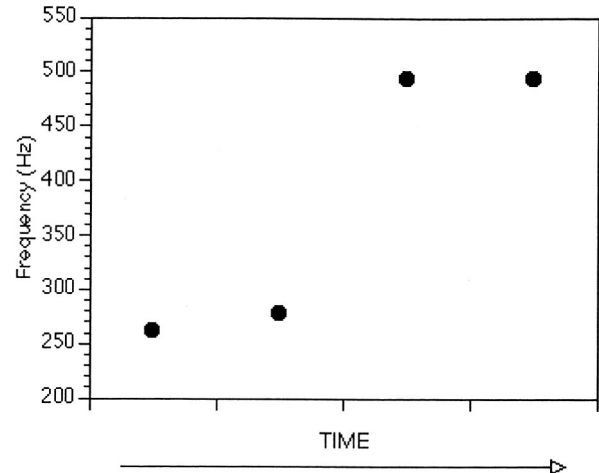
Procedure

The procedures were the same as that of Experiment 1A, the only differences relating to the nature of the stimulus elements, as described above. The timing of the stimuli, pauses, and prompts was identical to the timing in Experiment 1A.

Results

We performed the same statistical analyses as used in Experiment 1A. During the training phase, the Experiment 1B (visual) experimental group made correct match–mismatch decisions on 86% of the trials, whereas the Experiment 1C (auditory) experimental group scored 96%. We compared the training means across the three experiments, which revealed a main effect of modality, $F(2, 27) = 24.30$, $p < .0001$. Thus, auditory training performance was significantly better than visual performance ($p < .005$), which in turn was significantly better than tactile performance ($p < .001$). Because the training task essentially involves remembering and comparing sequences within pairs, the results may elucidate possible differences among the three modalities in representing and maintaining sequential information (Penney, 1989). It is also possible that these results instead are due to factors such as differential discriminability or perceptibility of sequence elements in different sensory domains.

Results for the test phase in Experiment 1B revealed that the control group correctly classified 47% of the test sequences, whereas the experimental group correctly classified 63% of the test sequences. An ANOVA (experimental vs. control group) indicated that the main effect of group was significant, $F(1, 18) = 3.15$, $p < .01$. Single group *t* tests revealed that the control group's performance was not significantly different from chance, $t(9) = -1.11$, $p = .3$, whereas the experimental group's performance was significantly different from chance, $t(9) = 3.03$, $p < .05$.

---

[6] This particular set of tones was used because it avoids familiar melodies (Dowling, 1991).

Results for the auditory (Experiment 1C) test phase revealed that the control group correctly classified 44% of the test sequences, whereas the experimental group correctly classified 75% of the test sequences. An ANOVA (experimental vs. control group) indicated that the main effect of group was significant, $F(1, 18) = 7.08$, $p < .001$. Single group t tests revealed that the control group's performance was marginally worse than chance, $t(9) = -2.25$, $p = .051$, indicating that our test stimuli were biased against a positive effect of learning. The experimental group's performance was significantly different from chance, $t(9) = 7.45$, $p < .001$.

Participants' verbal reports in Experiments 1B and 1C were similar to those in Experiment 1A. Namely, the most common report given was that participants were basing their classification decisions on how similar or familiar the sequences were relative to the training items. None of the participants was able to verbalize any of the rules governing the sequences. Therefore, it appears that participants generally did not benefit from explicit knowledge of the sequence structure.

These results indicate that both the visual and the auditory experimental groups significantly outperformed the control groups, with participants unable to verbalize how the legal and illegal sequences differed. Hence, participants appear to have implicitly learned aspects of the statistical structure of the visual and auditory input. These initial analyses suggest commonalities among tactile, visual, and auditory statistical learning.

However, one striking difference is that the auditory test performance was substantially better than tactile or visual performance (75% vs. 62% and 63%; see Figure 6). Submitting these three test performances to an ANOVA reveals a main effect of modality, $F(2, 27) = 3.43$, $p < .05$, with the effect due to the auditory performance being significantly better than both touch and vision ($ps < .05$). Thus, it appears that in this task, auditory statistical learning was more proficient than both tactile and visual learning. This is in accord with previous research emphasizing audition as being superior among the senses in regard to temporal processing tasks in general (e.g., Freides, 1974; Handel & Buffardi, 1969; Sherrick & Cholewiak, 1986).
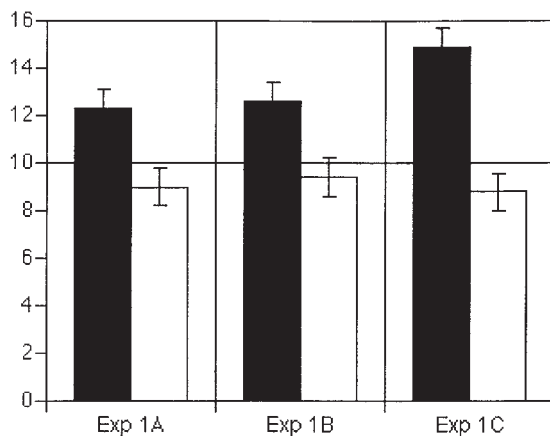

Figure 6. Experiment (Exp) 1: Mean number of correct test responses out of 20 (plus standard error) for the experimental (indicated by solid bars) and control (indicated by open bars) groups. Ten is the level expected for chance performance.

Discussion

The previous analyses have offered a quantitative comparison among tactile, visual, and auditory learning, revealing better learning in the auditory condition. One possible objection to this conclusion is that the auditory experiment differs from the first two experiments in that pitch, instead of space, is the primary stimulus dimension being manipulated. A different possibility would have been to set up five speakers at five different spatial locations, each one producing the same pitch stimulus at different times in the sequence, much like the visual stimuli were displayed in Experiment 1B. However, it has been proposed that for the auditory modality, pitch is, in a sense, equivalent to space (Kubovy, 1988). Shamma (2001) argued that the auditory nervous system transforms sound input, through the cochlea, into spatiotemporal response patterns, and therefore the visual and auditory systems process spatial and temporal input, respectively, in computationally similar ways. Thus, the perception of pitch and the perception of visual–spatial patterns may arise through similar computational algorithms in the two sensory modalities. For this reason, we believe that the most appropriate test for auditory statistical learning is to use stimulus elements that differ along the dimension of pitch rather than that of space. This is consistent with previous tests of auditory AGL to use stimulus elements that vary in terms of pitch or syllable rather than space. Although this research has found similar statistical learning performances in vision and audition (Fiser & Aslin, 2002; Saffran, 2002), our data suggest a quantitative advantage for auditory learning relative to tactile and visual learning.

We might also ask whether there were any *qualitative* learning differences among the three modalities. For example, were there particular test sequences within each modality that participants were better or worse at correctly endorsing? Which types of statistical information did participants within each modality rely on to perform the test task? To answer these questions, we present several additional analyses.

We first investigated whether certain sequences were easier or more difficult to classify for each modality. We conducted item analyses across the three sense modalities, entering the test performance data averaged across subjects for each sequence. This two-way ANOVA (Modality × Sequence) resulted in main effects of modality, $F(2, 540) = 4.73$, $p < .01$, and sequence, $F(19, 540) = 1.69$, $p < .05$, but no interaction of modality and sequence, $F(38, 540) = 1.20$, $p = .2$.

To get a better idea about which sources of information are most valuable for each modality, we analyzed each test sequence in terms of the information content that participants may have used to guide test performance. We used five fragment measures: associative chunk strength, novelty, novel fragment position (NFP), initial anchor strength (I-anchor), and final anchor strength (F-anchor). Associative chunk strength is calculated as the average frequency of occurrence of each test item's fragments (bigrams and trigrams), relative to the training items (Knowlton & Squire, 1994). Novelty is the number of fragments that did not appear in any training item (Redington & Chater, 1996). NFP is measured as the number of fragments that occur in novel absolute positions where they did not occur in any training item (Johnstone & Shanks, 1999). We designed the I-anchor and F-anchor measures to indicate the relative frequencies of initial and final fragments in similar positions in the

training items. Previous studies used a single anchor strength measure (e.g., Knowlton & Squire, 1994) instead of calculating the initial and final measures separately, as we do here. We consider I-anchor and F-anchor separately to determine whether modality constraints lead participants to be more or less sensitive to the beginnings or endings of sequences.[7] Finally, we used a measure of global similarity, which is the number of elements by which a test item is different from the nearest training item (Vokey & Brooks, 1992).

We computed these six measures for each of the 20 test sequences, and the results are listed in Table 1. Inspection of this table reveals that the legal and illegal test sequences differ considerably in terms of their chunk, I-anchor, F-anchor, novel, and NFP information. It is therefore likely that one or more of these information sources guided participants in making their classification judgments at test.

To see which information sources were used for each modality, we used regression analyses. Our initial regression model contained the six sources of information listed in Table 1 as predictors, in addition to two other predictors: length of each sequence, as measured by the number of elements per sequence, and legality, which was simply an index of whether the sequence was legal or illegal. Because these eight predictors are highly correlated with one another, we submitted them to a principal-components analysis (PCA) to reduce the number of predictors to use in the regression analyses. The results of the PCA revealed that the eight predictors could be reduced to two components, explaining 87.7% of the variance. These two components are listed in Table 2.

As can be seen, the first component is roughly a measure of chunk strength, including I-anchor and F-anchor, and is also an inverse measure of novelty and NFP. This is intuitive, because a sequence with a high chunk or anchor strength contains fewer novel fragments. The second component is nearly equivalent with length. With these results in mind, we decided to use three predictors in our multiple regression model: I-anchor, F-anchor, and length. Note that in essence, what we did was separate the first component (which is roughly equivalent to chunk strength) into initial and final chunk strength predictors. We did this with the expectation that the multiple regression analysis might reveal possible modality constraints related to beginning or ending sequence biases.

The results of the regression analyses will inform us as to which of these three measures best predict whether a participant in each

sensory condition will endorse a test sequence. We performed one linear regression for each modality. The results reveal that length ($p < .05$) and I-anchor ($p < .005$) were good predictors for tactile endorsements. F-anchor ($p < .005$) was a good predictor for auditory endorsements. None of the three predictors was a statistically significant predictor for visual endorsements.

In summary, the item analyses revealed no differences in terms of performance on individual sequences across the modalities. However, the multiple regression analyses revealed that there may be differences in terms of which sources of information are most important for test performance in each of the three modalities. We found that tactile learners were most sensitive to the length of the sequence and the fragment information at the beginning of a sequence, auditory learners were most sensitive to fragment information at the end of a sequence, and visual learners were biased toward neither the beginning nor the ending of the sequences. Thus, these preliminary analyses suggest that not only does auditory statistical learning of tone sequences have a quantitative advantage over tactile and visual learning, there also may be qualitative differences among the three modalities. Specifically, tactile learning appears to be sensitive to initial item chunk information, whereas auditory learning is most sensitive to final item chunk information.

Experiment 2: Tactile, Visual, and Auditory Statistical Learning

The first three experiments assessed statistical learning of tactile, visual, and auditory sequences. The results suggest the presence of modality differences affecting learning. Specifically, there was a *quantitative* learning difference in that auditory learning was superior to the other two senses. There was also evidence for *qualitative* learning differences in that the sense modalities appeared to be differentially sensitive to the initial or final aspects of the sequences. However, one unresolved question is whether the observed learning differences are merely the result of low-level, perceptual effects of the particular stimulus elements used in the three experiments. For example, it is possible that auditory learning was more effective because the set of tones used in Experiment 1C may have been more distinctive than the set of vibration pulses or visual stimuli used in Experiments 1A and 1B. Similarly, recall that auditory training performance was significantly better than visual or tactile performances; perhaps the superior auditory test scores were due to better performance in the training phase.

To better control for perceptual and training effects, we conducted Experiment 2, which was similar to the first set of experiments except for several crucial modifications. We used a pretraining phase to assess the perceptual comparability of the stimulus elements across modalities. Also, we used a modified training task in which participants observed a sequence followed by a bigram fragment and then judged whether the bigram fragment had occurred within the sequence. We adopted this new training task to ensure similar training performance levels across the three modalities. In addition, we used a randomized design to ensure that any differences across conditions were not the result of

Table 2
Results of Principal-Components Analysis

Measure	Component 1	Component 2
Chunk	.950	.179
Novel	−.953	−.002
NFP	−.945	.153
Sim.	−.696	.536
Length	−.154	.949
I-anchor	.903	.007
F-anchor	.846	.302
Legality	.947	.184

Note. NFP = Novel fragment position; Sim. = similarity; I-anchor = initial anchor strength; F-anchor = final anchor strength.

[7] Meulemans and Van der Linden (2003) also used separate I-anchor and F-anchor measures.

differences in population samples. Finally, we provided a more substantive test for qualitative learning differences by incorporating test stimuli that could better assess whether participants were differentially sensitive to statistical information in the beginnings or endings of sequences. Our hypothesis, following the analyses of Experiment 1, was that participants would be more sensitive to the initial fragments when exposed to tactile sequences, whereas they would be more sensitive to the final fragments when exposed to auditory sequences.

Method

Participants

An additional 48 undergraduates (8 in each condition) were recruited from introductory psychology classes at Cornell University.

Apparatus

The apparatus was the same as in Experiment 1.

Materials

To generate the stimuli used for Experiment 2, we created a new finite-state grammar (Figure 7). This grammar was created with two main constraints in mind. First, we intended it to be more complex than that used in Experiment 1. The new grammar can generate up to 75 sequences between three and seven elements in length (as opposed to 23 sequences in Experiment 1), allowing for a more difficult learning task. Second, we created the new finite-state grammar to allow us to test the hypothesis that learners are more or less sensitive to beginning or ending aspects of sequences in each sense modality. The grammar is symmetrical in terms of the number of possible bigrams and trigrams allowed in initial and final positions.[8] Thus, it is not biased toward the beginning or ending aspects of sequences in terms of the amount of chunk information available. This allows us to have better control over what parts of the sequences may be useful for the learner.

The five stimulus elements making up the sequences were identical to those used in Experiment 1 except for the auditory tones. The tone set used for the auditory stimuli was slightly different from before, consisting of 220 Hz, 246.9 Hz, 261.6 Hz, 277.2 Hz, and 329.6 Hz (i.e., the musical notes A, B, C, C#, and E, respectively). As with the previous tone set, we used these tones because they avoid familiar melodies (Dowling, 1991). Additionally, this new tone set spans a smaller frequency range (220 Hz to 329.6 Hz, as opposed to 261.6 Hz to 493.8 Hz).
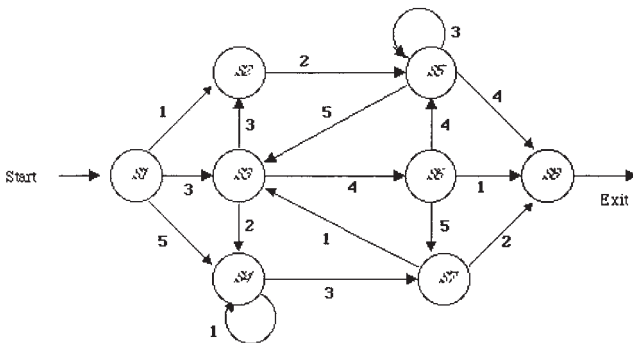

Figure 7. Artificial grammar used in Experiment 2. The numbers 1–5 correspond to each of the five possible stimulus elements for the tactile, visual, and auditory modalities (depending on the experimental condition). S = State.

We also tested all materials for their discriminability across modalities. Ten separate participants took part in a discrimination task in which they received two stimuli (within the same modality) and judged whether they were the same or different. Participants were presented with all of the possible pairwise combinations for each modality. The data revealed that participants were able to correctly discriminate the stimuli at near-perfect levels across all three modalities (tactile: 95%; visual: 98.3%; auditory: 98.8%), with no statistical difference in performance among modalities ($p = .87$).

Pretraining phase. For the pretraining phase, each of the five stimulus elements was paired with each other to give every possible combination ($5^2 = 25$ possible combinations). Because responses for pairs such as 3–2/2–3 and 1–4/4–1 were averaged together in the analysis (see the *Results* section), we presented the 5 pairs that contain identical elements two times instead of once (e.g., 1–1, 2–2). This gave a total of 30 stimulus pairs. Each stimulus element had a duration of 250 ms, and elements were separated by 250 ms. The pretraining materials are listed in Appendix B.

Training phase. A total of 24 legal sequences were generated from the new finite-state grammar and used for the training phase. Each of these sequences was coupled with a particular bigram fragment. For half of the sequences, the bigram appeared within the sequence (e.g., 3–4–5–1–2–3–2 and 1–2). For the other half of the sequences, the bigram itself did not occur within the sequence, but the elements composing the bigram did (e.g., 1–2–3–5–2–3–2 and 1–3). In all cases, the bigrams presented after the sequence were legal according to the finite-state grammar. Each stimulus element had a duration of 250 ms and was separated from the elements before and after by 250 ms. A 2-s pause separated the sequence from the bigram. The training materials are listed in Appendix C.

Test phase. The test set consisted of 16 novel legal and 16 novel illegal sequences. Legal sequences were produced from the finite-state grammar in the normal fashion. We produced illegal sequences by changing two elements of each legal test sequence. We created 8 of the illegal sequences, referred to as illegal–initial sequences, by modifying the second and third elements of a legal sequence (e.g., legal: 5–1–3–1–4–5–2; illegal: 5–5–2–1–4–5–2). We created the other 8 illegal sequences, referred to as illegal–final sequences, by modifying the third-to-last and second-to-last elements of a legal sequence (e.g., legal: 3–2–3–1–2–3–2; illegal: 3–2–3–1–5–2–2). Each illegal sequence was paired with the legal sequence from which it was generated, counterbalanced so that all sequences appeared both first and last, giving a total of 32 test pairs. Each stimulus element had a duration of 250 ms and was separated by 250 ms. A 2-s pause separated one sequence from the next within a pair. Table 3 lists the test materials.

We created the Experiment 2 test sequences so that information about legal element repetitions would not be useful. For instance, Table 3 reveals that out of the 32 test sequences, 18 are relevant for element repetitions, and the other 14 sequences are neutral in regard to element repetition information. If one uses the strategy of choosing the sequence within a pair containing legal element repetitions (i.e., those repetitions seen in the training sequences), this would lead to only 8 out of 18 correct endorsements. Thus, such a strategy is actually worse than random guessing, meaning that the test sequences are well controlled in terms of element repetition information.

Additionally, as we did in Experiment 1, we can analyze the test sequences in terms of chunk, novelty, and similarity information in relation to the training set. We divided the test set into four groups: legal–initial, illegal–initial, legal–final, and illegal–final. We then analyzed each group in terms of the fragment measures and made statistical comparisons among the various groups.

Table 3 shows the associative chunk strength, I-anchor, F-anchor, novelty, NFP, and similarity measures for each of these four groups.

---

[8] There are 6 unique initial bigrams, 6 unique final bigrams, 13 unique initial trigrams, and 13 unique final trigrams.

Table 3
Fragment Measures for Experiment 2 Test Sequences

Item	Chunk	Novel	NFP	Sim.	I-anchor	F-anchor
Legal–initial sequences						
5–1–3–1–4–5–2	6.27	0	0	2	2.50	5.00
1–2–3–3–5–5–4–1	3.27	2	8	3	2.00	2.00
3–3–2–3–3–4	5.56	1	3	2	2.00	2.00
1–2–5–3–2–3–4	4.91	3	5	3	1.00	2.00
3–4–5–1–3–2–4	6.09	0	1	2	6.00	1.00
1–2–5–4–4–3–4	4.18	1	5	3	1.00	2.00
1–2–5–4–4–4	5.56	1	5	3	1.00	5.00
1–2–5–3–2–4	3.22	3	7	4	1.00	1.00
Average	4.88	1.38	4.25	2.75	2.06	2.50
Illegal–initial sequences						
5–5–2–1–4–5–2	4.00	4	6	3	0.00	5.00
1–1–1–3–5–4–1	2.73	1	9	5	0.00	2.00
3–5–5–3–3–4	2.44	4	6	2	0.00	2.00
1–4–4–3–2–3–4	6.73	1	3	3	0.00	2.00
3–1–2–1–3–2–4	4.09	2	6	2	0.00	1.00
1–1–1–4–4–3–4	4.91	1	3	2	0.00	2.00
1–5–1–4–4–4	6.44	2	4	2	0.00	5.00
1–4–2–3–2–4	4.67	3	8	3	0.00	1.00
Average	4.50	2.25	5.62	2.75	0.00	2.50
Legal–final sequences						
1–2–3–3–3–4	4.33	2	3	3	2.00	2.00
3–2–3–1–2–3–2	7.82	0	0	2	3.50	6.00
3–3–2–3–3–3–4	4.82	2	6	3	2.00	2.00
3–4–4–3–3–3–4	4.82	1	3	3	6.00	2.00
3–2–1–3–1–4–1	5.00	1	7	4	2.50	2.50
3–4–4–3–5–4–1	4.18	1	6	3	6.00	2.00
3–2–3–1–4–1	7.00	0	0	2	3.50	2.50
1–2–3–5–3–2–4	4.27	2	3	2	2.00	1.00
Average	5.28	1.12	3.50	2.75	3.44	2.50
Illegal–final sequences						
1–2–3–1–1–4	5.22	2	3	3	2.00	0.00
3–2–3–1–5–2–2	4.55	5	5	2	3.50	0.00
3–3–2–3–5–4–4	5.64	0	7	4	2.00	2.50
3–4–4–3–5–5–4	3.73	4	6	3	6.00	0.00
3–2–1–3–4–3–1	4.18	4	6	4	2.50	0.00
3–4–4–3–2–1–1	5.09	1	5	4	6.00	0.00
3–2–3–5–1–1	5.11	1	4	3	3.50	0.00
1–2–3–5–4–1–4	4.00	1	2	1	2.00	0.00
Average	4.69	2.25	4.75	3.00	3.44	0.31

Note. NFP = Novel fragment position; Sim. = similarity; I-anchor = initial anchor strength; F-anchor = final anchor strength.

Legal–initial and illegal–initial items differed only in terms of I-anchor (2.06 vs. 0.00, $p < .05$). Likewise, legal–final and illegal–final items differed only in terms of F-anchor (2.50 vs. 0.31, $p < .05$). Legal–initial and legal–final items were statistically identical across all measures ($ps > .2$). Illegal–initial and illegal–final items differed in terms of both I-anchor (0.00 vs. 3.44, $p < .001$) and F-anchor (2.50 vs. 0.31, $p < .05$). Thus, in terms of fragment information, the only differences among the four groups of test sequences lies among the dimensions of initial and final chunk anchor strengths. This means that we can clearly examine differences in participants' sensitivities to initial and final fragment information across the three sensory modalities.

Procedure

The overall procedure was similar to that of the previous experiments but included an extra pretraining phase as well as a modified training task. Participants were randomly assigned to one of six conditions: tactile, visual, auditory, tactile control, visual control, or auditory control. The three control conditions were identical to their respective experimental conditions except that the controls participated in the pretraining and test phases only.

All participants in the tactile conditions were assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) to determine their preferred hand.

Pretraining phase. As already described, a separate group of participants had participated in a simple discrimination task, which revealed that the stimuli are easily discriminable across the modalities. To provide an additional test of perceptual comparability, we incorporated the pretraining phase into the current experiment. As an additional benefit, this procedure also served to familiarize participants with the actual stimulus elements before they were exposed to the training sequences.

Participants were informed that they would observe two stimuli, one following the other. The stimuli consisted of vibration pulses, visual stimuli, or tones, depending on the experimental condition. Participants were required to judge how similar the two stimuli were to each other and give a rating between 1 and 7, where 1 corresponded to most dissimilar and 7 to most similar. Participants in the tactile conditions were told to base their ratings on the vibration pulses' proximity to each other, as all vibration pulses were identical except for which fingers were stimulated. Similarly, participants in the visual conditions also were told to base their ratings on the stimuli's proximity, as the stimuli themselves were identical and differed only in terms of where they were located. Participants in the auditory conditions were told to base their ratings on the pitches of the tones.

Before the rating task began, participants were exposed to each of the five possible stimuli, one at a time, so that they knew what the possible stimuli were. Then they were presented with each of the 30 possible pairs listed in Appendix B, in random order for each participant. All stimuli were delivered for a duration of 250 ms with a 250 ms pause occurring between the stimuli within a pair. A prompt containing a reminder of the rating scheme appeared on the screen, and the participant used the keyboard to give a numerical response between 1 and 7. Following a 2-s pause after the rating was given, the next stimulus pair was delivered.

Training phase. As in Experiment 1, the purpose of the training phase was for the participants to attend to the legal training sequences without explicit instruction that the sequences contained statistical regularities. On the basis of pilot studies, we modified the training procedure slightly from Experiment 1 in an attempt to equate training performance across the three modalities.

At the beginning of the training phase, participants were instructed that they would observe a particular sequence of stimuli and then, after a slight pause, would observe two additional elements. The task was to decide whether the pair of elements had occurred within the sequence in the same order and then to press the appropriate key, *Y* for yes, *N* for no. The training sequence–pair combinations from Appendix C were presented in random order for three blocks, for a total of 72 training trials. Stimulus elements had a duration of 250 ms and were separated by 250-ms pauses. A 2-s pause occurred between each sequence and each pair of elements. One second after the last element of the stimulus pair occurred, a prompt was displayed on the screen asking for the participant's response. The next sequence–pair combination began after a 2-s pause.

Test phase. The purpose of the test phase was to assess how well participants learned the statistical regularities of the training set and could generalize such knowledge to novel stimuli in a classification task. At the beginning of the test phase, participants were instructed that all of the sequences they had been exposed to in the previous phase of the experiment were generated by a complex set of rules. They now would be exposed to new sequences, presented in groups of two. One of the se-

quences in each pair was generated by the same rules as before, whereas the other was not. The participants' task was to choose which sequence was generated from the same rules by pressing a key marked *1* or *2*, signifying the first or second sequence, respectively. The test sequence pairs from Table 3 were presented in random order for each participant. Stimulus elements had a duration of 250 ms and were separated by 250-ms pauses. A 2-s pause occurred between the two sequences of a pair. One second after the second sequence occurred, a prompt was displayed on the screen asking for the participant's response. The next pair of sequences began following a 2-s pause.

Results

Pretraining Phase

We collected similarity ratings from all participants and averaged them for each element pair combination to form three similarity matrices, one for each modality. Within each modality, the experimental and control ratings were combined. We submitted each similarity matrix to a multidimensional scaling (MDS) procedure (euclidean model) using SPSS 10.0 for Windows. A one-dimensional solution provided a good fit for each of the three modalities, with stress values less than 0.0500 in all cases (touch = 0.0056; vision = 0.0086; audition = 0.0470). As can be seen from Figure 8, the tactile and visual solutions contain clearly separated stimuli, in accord with the linear relationship of the actual stimuli (e.g., the tactile vibration elements are in the expected order, beginning with the thumb pulse, then the second finger, then third finger). In slight contrast, the auditory solution contains two tones, the third (261.6 Hz) and fourth (277.2 Hz), that are clustered together in state space.

We interpret these MDS solutions as depicting that, overall, the stimuli in the three modalities are perceived in psychologically similar ways. The only noticeable difference is that two of the tones may have similar perceptual representations.

Training Phase

The mean training performance out of 72 for each modality was 43.38 (60.3%) for tactile, 50.13 (69.6%) for visual, and 48.25
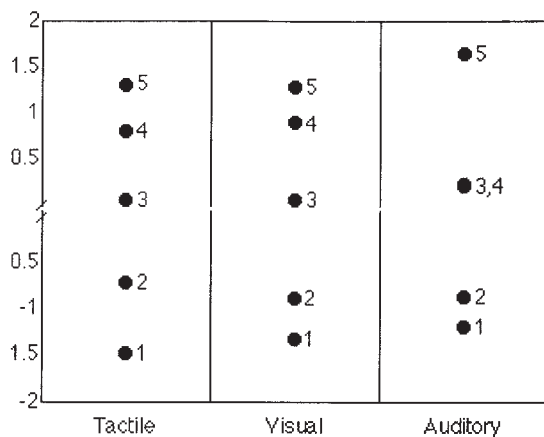

Figure 8. Multidimensional scaling solutions using the pretraining data for each of the three sensory modalities. The numbers correspond to each of the five possible stimulus elements for each modality.

(67.0%) for auditory. The data were submitted to a one-way ANOVA with the factor of modality. There was a marginally significant main effect, $F(2, 21) = 2.81$, $p = .083$. Post hoc tests revealed a significant difference between the tactile and visual training means ($p < .05$). These results indicate that although the tactile training performance was somewhat lower than the auditory and visual performances, in general, scores across the three modalities were roughly equivalent.

Test Phase

The mean test scores out of 32 for each group were 15.75 for tactile–control (49.2%), 16.00 for tactile–experimental (50.0%), 14.38 for visual–control (44.9%), 15.88 for visual–experimental (49.6%), 15.25 for auditory–control (47.6%), and 19.25 for auditory–experimental (60.2%). Figure 9 displays Experiment 2 test performance.

Recall that the control groups participated in the pretraining and test phases only. Therefore, the best way to assess learning within each sensory modality is to compare experimental group performance with both hypothetical chance levels (i.e., 50%) and control group performance (see Redington & Chater, 1996). Individual single group *t* tests comparing experimental group performance with theoretical chance reveal that only the auditory group performed significantly better than chance, $t(7) = 3.61$, $p < .01$. Likewise, paired *t* tests reveal that only the auditory experimental group performed better than its control group, $t(7) = 3.86$, $p < .01$. These results reveal that only the auditory experimental group learned the statistical regularities of the training corpus.

We next submitted the experimental group data to a repeated-measures ANOVA with the between-subjects factor of modality and the within-subject factor of initial versus final group sequences. We found a main effect of modality, $F(2, 21) = 4.95$, $p < .05$, a main effect of initial–final group, $F(2, 21) = 8.61$, $p < .01$, and no interaction ($p = .25$). The effect of initial–final group arises because test performance across modalities was generally better at initial group sequences. The main effect of modality arises because the auditory group was significantly better than both the tactile and the visual groups ($p s < .05$). Thus, these analyses confirm the presence of modality differences in learners' test performances.

The ANOVA revealed no interaction between modality and initial versus final group sequences. However, because Experiment 1 suggested the presence of such modality differences, we continued to explore possible differences with planned comparisons. As described earlier, the test sequences were created such that test pairs differed only in terms of I-anchor and F-anchor measures. Thus, we can easily determine whether each sensory modality was better at discriminating sequences on the basis of initial or final fragment information. For the experimental group test performance in each modality, we considered initial and final test pairs separately. Performance on the 16 initial test pairs (i.e., test pairs consisting of one legal–initial and one illegal–initial sequence) was 8.50 (53.1%) for touch, 9.50 (59.4%) for vision, and 10.13 (63.3%) for audition. We conducted a one-way ANOVA that revealed no statistical differences among the three modalities, $F(2, 21) = 1.69$, $p = .21$.

Next, we consider performance on the 16 final test pairs for the experimental conditions (i.e., test pairs consisting of one legal–
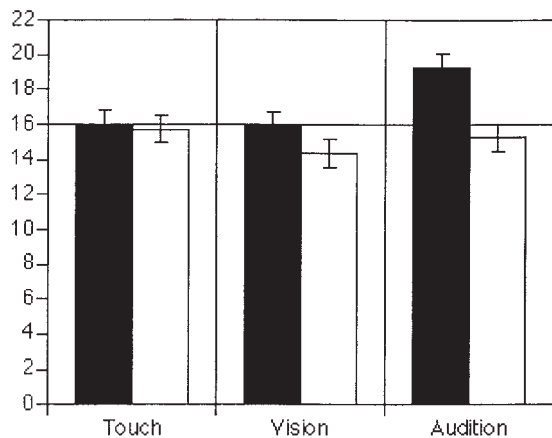
Figure 9. Experiment 2: Mean number of correct test responses out of 32 (plus standard error) for the experimental (solid bars) and control (open bars) groups. Sixteen is the level expected for chance performance.

final and one illegal–final sequence). Performance was 7.50 (46.9%) for tactile, 6.38 (39.8%) for visual, and 9.13 (57.0%) for auditory. An ANOVA revealed a main effect of modality, $F(2, 21) = 3.98$, $p < .05$. Specifically, auditory performance was significantly better than visual performance ($p < .05$).

To summarize, the auditory group alone showed learning. Furthermore, the auditory superiority appears to be largely due to better performance, relative to vision and touch, on fragment–final test pairs. Thus, consistent with Experiment 1, Experiment 2 provides evidence for both quantitative and qualitative learning differences across the senses.

Discussion

Experiment 2 more closely examined modality differences in statistical learning by attempting to control for two variables that could have influenced performance in Experiment 1: low-level perceptual factors and training performance effects. We controlled the first variable by introducing a pretraining phase, in which participants observed all combinations of element pairs and gave similarity ratings for each. These ratings were then submitted to an MDS analysis, and the results provide good indication that, across modalities, the stimuli were represented similarly. Furthermore, an additional control task revealed that the stimuli within each modality were easily discriminated.

Even if the stimuli across all three conditions were comparable in terms of their perceptibility, it is still possible that uneven training performances could lead to differences in test performance. For example, in Experiment 1, the auditory group's training performance was significantly better than both visual and tactile training performance. The improved training performance may have led to better encoding of the relevant fragment information, resulting in a better ability to classify novel sequences at test. For Experiment 2, we controlled training performance effects by using a training task that resulted in relatively comparable training scores. This task proved more difficult than that used in the first set of experiments, with scores ranging between 60% and 69%. Auditory training performance was equivalent to visual performance and only slightly better than tactile performance.

Even after we controlled for the perceptual and training effects, auditory test performance still was significantly better than tactile and visual performance. In fact, only the auditory group showed a main effect of learning. Experiment 2 also shows that the auditory modality's better performance was due to a heightened sensitivity to sequence–final fragment information.

Comparing the results of Experiment 2 with those of Experiment 1 reveals strong similarities. Both experiments showed an auditory advantage for classifying novel sequences in regard to their legality. Both experiments also revealed differences in terms of whether initial or final fragment information was more important for each modality. The regression analyses in the first set of experiments suggested a tactile–initial and an auditory–final effect. The auditory–final effect was confirmed in Experiment 2.

One potentially troubling aspect of Experiment 2 is that there appeared to be no tactile learning. This may be a result of the more complex grammar and more difficult test that we used in Experiment 2. Tactile memory for serial lists may be weaker compared with vision and audition (Mahrer & Miles, 1999) and, if so, may hinder learning during the training phase. The lack of learning leaves open the possibility that the tactile learning we found in Experiment 1A was spurious. To replicate the tactile learning from Experiment 1A, we conducted an additional tactile learning experiment.[9]

We hypothesized that using a slightly modified training scheme, one that is less computationally demanding, might allow participants to better encode the training regularities and lead to successful learning. Instead of a training sequence being followed by a pair of elements (i.e., a bigram), each training sequence was followed by a single element. The participants' task was to judge whether the single element had been a part of the preceding sequence. In all other respects this additional experiment was identical to the tactile condition in Experiment 2. On the training task, participants scored 62.2 out of 72 (86.4%), and on the test task they scored 18.63 out of 32 (58.2%). An ANOVA (experimental vs. control) comparing the test performance with the control performance revealed a significant effect, $F(1, 14) = 5.83$, $p < .05$. Thus, under a slightly less complex training condition that allowed better training performance, tactile participants showed a main effect of learning in the test phase. This confirms the tactile learning we found in Experiment 1.

Finally, it could be argued that the superior performance in the auditory condition was due to the presence of relative pitch information present in the tone sequences (e.g., see Saffran & Griepentrog, 2001). Because the auditory sequences can be construed as melodies, they contain information about not just absolute pitch but also relative pitch in the form of familiar musical intervals (e.g., perfect fourths and major sevenths). To eliminate this information and provide an auditory task more comparable to the other experiments, we conducted a new auditory experiment identical to Experiment 2 but with a different tone set. The new tones were created so that they would neither conform to standard musical notes nor contain standard musical intervals between them. The

---

[9] Although there also was no effect of learning in Experiment 2's visual condition, we did not feel it was necessary to conduct an additional visual learning experiment because visual statistical learning has been demonstrated in previous research (e.g., Fiser & Aslin, 2002).

tones, having frequencies of 210 Hz, 245 Hz, 286 Hz, 333 Hz, and 389 Hz, were equally spaced in log frequency, on the basis of a ratio of 7/6 (1.166) and its powers 2, 3, and 4 (1.36, 1.59, 1.85, respectively). Of these ratios, only one comes close to a musical interval, minor 6 (ratio 1.6), which is not a very prominent interval. With this new tone set, participants scored 41.8 out of 72 (58.1%) on the training task and 18.7 out of 32 (58.4%) at test. The test scores were significantly greater than chance, $t(9) = 3.69$, $p < .01$, revealing an effect of learning. We also compared the test scores with the original Experiment 2 auditory scores and found no difference, $t(7) = -0.35$, $p = .74$. Thus, this additional experiment indicates that the presence or absence of musical interval information does not appear to affect learning and therefore is not the underlying cause for the superior auditory performance.

General Discussion

Our experiments provided a controlled investigation into the nature of statistical learning across the three primary sensory modalities. Unlike in previous statistical learning forays, we used comparable materials and identical procedures across the three senses. Additionally, in Experiment 2, we controlled for low-level perceptual as well as training performance effects, which allowed us to make direct comparisons among the modalities. In this section we discuss the main findings of these experiments and conclude by considering the underlying neural and cognitive mechanisms.

The first important finding is that touch can mediate statistical learning of sequential input. Experiment 1A revealed that after very brief (10 min) exposure to training sequences produced from an artificial grammar, participants were able to classify novel sequences as being either generated or not generated from the same grammar. To our knowledge, no other studies have demonstrated such a tactile learning ability, perhaps because of the belief that deep structure learning is beyond the tactile sense's capacity. Participants did not perform above chance or control level performance in Experiment 2, likely because the grammar was too complex and the distinction between legal and illegal sequences more subtle. However, when we used a slightly less demanding training scheme, participants displayed learning, which confirms that touch can mediate the learning of an artificial grammar.

The tactile learning in both experiments occurred in the absence of feedback and apparently without participants' awareness. Additionally, participants learned more than simply element frequencies, sequence lengths, or beginning and ending element legality, as these factors were controlled; rather, the legal and illegal test sequences differed in terms of adjacent element statistics. The learning process observed here in the tactile domain is conceptually similar to statistical learning found in the visual and auditory modalities (e.g., Fiser & Aslin, 2002; Saffran et al., 1999). Besides having theoretical importance, these data may have practical implications in the realm of communication devices for the visually or auditorily impaired, by capitalizing on people's ability to encode and represent tactile temporal order statistics.

The second important finding is the presence of modality constraints affecting statistical learning. These modality constraints took two main forms: a quantitative effect, and a qualitative effect. The quantitative effect was evidenced by the fact that auditory statistical learning was better than both tactile and visual learning.

The auditory advantage occurred even after we controlled for training performance and stimulus element perceptibility across the modalities and eliminated standard musical intervals from the tone sequences. Our results thus suggest that the modality constraints affected learning itself rather than a lower level perceptual process that might have in turn affected learning performance. Although previous research has pointed to an auditory advantage for low-level temporal processing tasks (Mahar et al., 1994; Sherrick & Cholewiak, 1986), our results appear to be some of the first evidence that such an advantage extends to more complex processes, namely statistical learning of sequential input.

In addition to the quantitative effect was a qualitative learning effect that took the form of biases regarding which aspects of a sequence each sensory system is more or less attuned toward. We found evidence that learners of tactile sequences may have been more sensitive to fragment–initial information, whereas learners of auditory material were more sensitive to fragment–final information. These biases suggest that each sensory system may apply slightly different computational strategies when processing sequential input. The auditory–final bias is interesting because it mirrors the modality effect in serial recall, in which a more pronounced recency effect (i.e., greater memory for items at the end of a list) is obtained with spoken as compared with written lists (e.g., Crowder, 1986). This may indicate that similar constraints affect both explicit encoding of serial material and implicit learning of statistical structure. In both cases, learners appear to be more sensitive to material at the end of sequences or lists for auditory input. It may prove fruitful to further explore this hypothetical connection between serial list learning and implicit sequential learning; the results may inform research in both domains, which have traditionally remained separate endeavors.

In contrast to the quantitative and qualitative effects we found, two previous studies comparing auditory and visual statistical learning did not report any modality differences. Fiser and Aslin (2002) found human visual statistical learning of spatiotemporal sequences to be very similar to learning in the auditory domain with temporal sequences (e.g., Aslin, Saffran, & Newport, 1998) and in the visual domain with spatial structures (e.g., Fiser & Aslin, 2001). Similarly, Kirkham et al. (2002) concluded that infant visual statistical learning was similar to auditory learning (e.g., Saffran et al., 1996). However, Fiser and Aslin's (2002) and Kirkham et al.'s (2002) conclusions were based on studies that did not use comparable stimuli or procedures across the modalities. Thus, neither of these two studies are adequate for making fine-tuned cross-modal comparisons. We agree that there are similarities in how infants and adults learn sequential patterns across vision and audition, but our data reveal important differences, as well. Saffran (2002) used more comparable procedures and stimuli in her visual and auditory AGL experiments and found comparable overall learning performances across modalities. However, a subtle modality difference was revealed, showing that auditory learners were more sensitive to the presence or absence of predictive dependencies in the grammar, perhaps because they had more experience in the auditory domain for tracking such sequential structure.

Similarly, it could be argued that the auditory learning advantage we observed was merely due to participants having more prior experience listening to tone sequences compared with feeling vibration pulses or tracking visual spatiotemporal patterns. How-

ever, if this was the case, one would expect the auditory training performance levels to be substantially better than visual or tactile scores. In fact, auditory training performance was lower than visual performance in Experiment 2. Additionally, it is difficult to see how previous training exposure would lead to the qualitative differences we observed in terms of each sensory modality being differentially biased toward the beginning or ending of sequences.

If, as we suggest, modality constraints exist, what might be their nature? The answer clearly depends on how one views statistical or implicit learning itself. Because most researchers have tended to emphasize the similarities existing between statistical learning in various domains, this has led to implicit assumptions—or, in some cases, explicit statements (Kirkham et al., 2002)—depicting statistical learning as a single, domain-general mechanism. If this view is adopted, modality constraints are seen as influencing the processing of input sequences before the information is funneled into the presumed amodal statistical learning mechanism, allowing some types of input to be processed more or less efficiently.

Another view of statistical learning is that it is made up of various subsystems, each operating over different types of input and subserved by different brain areas. This view is supported by increasing evidence that unimodal brain areas contribute to the learning of statistical patterns. For example, in a functional magnetic resonance imaging study (Reber, Stark, & Squire, 1998), subjects learned to categorize visual patterns in terms of whether the patterns were similar to a previously seen corpus. It was found that occipital cortex (V1, V2) was instrumental for learning the structural regularities of the patterns, apart from being involved in visual perception itself. Similarly, Molchan, Sunderland, McIntosh, Herscovitch, and Schreuers (1994) found the primary auditory cortex to be involved in auditory associative learning. These studies indicate that relatively low-level unimodal sensory areas play an important role in learning environmental statistical regularities. It is our view that implicit statistical learning may be akin to perceptual priming, where modality-specific brain areas mediate learning by becoming more fluent at processing previously observed stimuli and/or stimuli that contain similar statistical properties as those viewed earlier (also see Chang & Knowlton, 2004). Within this framework, modality constraints may reflect general processing differences that exist among the various statistical learning subsystems, with the auditory system excelling at encoding statistical relations among temporal elements and the visual system specializing primarily in computing spatial relationships.

In conclusion, we have presented new evidence revealing both similarities and differences in statistical learning across the senses. An important target for future research is to uncover to what extent these modality constraints are related to previous reports of modality differences in perception and cognition. We anticipate that future studies, involving a combination of cognitive and neurophysiological methods, will further illuminate the nature of modality-constrained statistical learning across the senses.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321–324.

Beaman, C. P. (2002). Inverting the modality effect in serial recall. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 371–389.

Chang, G. Y., & Knowlton, B. J. (2004). Visual feature learning in artificial grammar classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 714–722.

Cleeremans, A. (1993). *Mechanisms of implicit learning: A connectionist model of sequence processing.* Cambridge, MA: MIT Press.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25,* 257–271.

Collier, G. L., & Logan, G. (2000). Modality differences in short-term memory for rhythms. *Memory and Cognition, 28,* 529–538.

Conway, C. M., & Christiansen, M. H. (2001). Sequential learning in non-human primates. *Trends in Cognitive Sciences, 5,* 539–546.

Craig, J. C., & Rollman, G. B. (1999). Somesthesis. *Annual Review of Psychology, 50,* 305–331.

Crowder, R. G. (1986). Auditory and temporal factors in the modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 2,* 268–278.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception and Psychophysics, 5,* 365–373.

Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics, 50,* 305–313.

Engle, R. W., & Mobley, L. A. (1976). The modality effect: What happens in long-term memory? *Journal of Verbal Learning and Verbal Behavior, 15,* 519–527.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499–504.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 458–467.

Freides, D. (1974). Human information processing and sensory modality: Cross-modal functions, information complexity, memory, and deficit. *Psychological Bulletin, 81,* 284–310.

Garner, W. R., & Gottwald, R. L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology, 20,* 97–109.

Gescheider, G. A. (1966). Resolving of successive clicks by the ears and skin. *Journal of Experimental Psychology, 71,* 378–381.

Gescheider, G. A. (1967). Auditory and cutaneous perceived successiveness. *Journal of Experimental Psychology, 73,* 179–186.

Glenberg, A. M., & Fernandez, A. (1988). Evidence for auditory temporal distinctiveness: Modality effects in order and frequency judgments. *Journal of Experimental Psychology, 14,* 728–739.

Glenberg, A. M., & Jona, M. (1991). Temporal coding in rhythm tasks revealed by modality effects. *Memory & Cognition, 19,* 514–522.

Glenberg, A. M., & Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 3–15.

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

Gomez, R. L., & Gerken, L. A. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Handel, S., & Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology, 21,* 256–266.

Hauser, M. D., Newport, E., & Aslin, R. N. (2001). Segmentation of the speech stream in a non-human primate: Statistical learning in cotton-top tamarins. *Cognition, 78,* B53–B64.

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.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain-general learning mechanism. *Cognition, 83,* B35–B42.

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.

Kubovy, M. (1988). Should we resist the seductiveness of the space:time::vision:audition analogy? *Journal of Experimental Psychology: Human Perception and Performance, 14,* 318–320.

Lechelt, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisited. *British Journal of Psychology, 66,* 101–108.

Mahar, D., Mackenzie, B., & McNicol, D. (1994). Modality-specific differences in the processing of spatially, temporally, and spatiotemporally distributed information. *Perception, 23,* 1369–1386.

Mahrer, P., & Miles, C. (1999). Memorial and strategic determinants of tactile recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 630–643.

Manning, S. K., Pasquali, P. E., & Smith, C. A. (1975). Effects of visual and tactual stimulus presentation on learning two-choice patterned and semi-random sequences. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 736–744.

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.

Meulemans, T., & Van der Linden, M. (2003). Implicit learning of complex information in amnesia. *Brain and Cognition, 52,* 250–257.

Molchan, S. E., Sunderland, T., McIntosh, A. R., Herscovitch, P., & Schreuers, B. G. (1994). A functional anatomical study of associative learning in humans. *Proceedings of the National Academy of Sciences, USA, 91,* 8122–8126.

Oatley, K., Robertson, A., & Scanlan, P. M. (1969). Judging the order of visual stimuli. *Quarterly Journal of Experimental Psychology, 21,* 172–179.

Oldfield, R. L. (1971). The assessment of handedness: The Edinburgh Inventory. *Neuropsychologia, 9,* 97–113.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401–426.

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition, 17,* 398–422.

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

Pothos, E. M., & Bailey, T. M. (2000). The role of similarity in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 847–862.

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior, 6,* 855–863.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious.* Oxford, England: Oxford University Press.

Reber, P. J., Stark, C. E. L., & Squire, L. R (1998). Cortical areas supporting category learning identified using functional MRI. *Proceedings of the National Academy of Science, 95,* 747–750.

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

Rubinstein, L., & Gruenberg, E. M. (1971). Intramodal and crossmodal sensory transfer of visual and auditory temporal patterns. *Perception and Psychophysics, 9,* 385–390.

Saffran, J. R. (2000). Non-linguistic constraints on the acquisition of phrase structure. In L. R. Gleitman & A. K. Joshi (Eds.), *Proceedings of the twenty-second annual conference of the Cognitive Science Society* (pp. 417–421). Hillsdale, NJ: Erlbaum.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language, 47,* 172–196.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology, 37,* 74–85.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70,* 27–52.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Seger, C. A. (1994). Implicit learning. *Psychological Bulletin, 115,* 163–196.

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends in Cognitive Sciences, 5,* 340–348.

Sherrick, C. E., & Cholewiak, R. W. (1986). Cutaneous sensitivity. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance, Vol. I: Sensory processes and perception* (pp. 1–12, 58). New York: Wiley.

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.

Wright, A. A. (2002). Monkey visual and auditory memory. In S. B Fountain, M. D. Bunsey, J. H. Danks, & M. K. McBeath (Eds.), *Animal cognition and sequential behavior: Behavioral, biological, and computational perspectives* (pp. 69–90). Boston: Kluwer Academic.

Appendix A

Training Stimuli for Experiment 1

Matched pairs	Mismatched pairs
1–2–1–1–1–3/1–2–1–1–1–3	1–2–3–5–2–5/1–2–3–5–2–3
4–1–1–3–5–2/4–1–1–3–5–2	1–2–3–5–2–3/1–2–3–5–2–5
4–1–3–5–2/4–1–3–5–2	4–3–5–2–3/4–3–5–2–5
1–2–5–5–5/1–2–5–5–5	4–3–5–2–5/4–3–5–2–3
4–1–3/4–1–3	1–2–5–5/1–2–1–3
1–2–3/1–2–3	1–2–1–3/1–2–5–5

Note. The numbers refer to a particular finger vibration (Experiment 1A), visual stimulus (Experiment 1B), or auditory tone (Experiment 1C).

Appendix B

Pretraining Materials for Experiment 2

1–1 (×2)	1–2	1–3	1–4	1–5
2–1	2–2 (×2)	2–3	2–4	2–5
3–1	3–2	3–3 (×2)	3–4	3–5
4–1	4–2	4–3	4–4 (×2)	4–5
5–1	5–2	5–3	5–4	5–5 (×2)

Note. ×2 indicates that this element pair was presented twice. The numbers 1–5 correspond to each of the five possible stimulus elements for the tactile, visual, and auditory modalities (depending on the experimental condition).

Appendix C

Training Materials for Experiment 2

5–3–1–3–2–3–4/5–3	3–4–4–5–4–4–4/3–5	5–1–1–1–3–2/3–2	1–2–3–5–2–3–2/1–3
3–2–3–1–4–5–2/2–3	3–2–3–1–3–2–4/3–4	3–4–5–1–4–4–4/3–4	5–3–1–4–4–3–4/4–5
1–2–3–5–4–4/3–5	3–4–4–5–4–5–2/5–3	5–3–1–4–5–2/3–1	3–4–5–1–4–1/4–3
3–4–5–1–4–5–2/1–4	5–1–3–1–2–3–2/2–1	3–4–4–5–4–1/4–5	5–1–3–1–4–1/4–3
3–4–5–1–2–3–2/1–2	3–3–2–5–4–5–2/4–3	5–3–1–4–4–4/4–4	3–4–4–3–3–4/3–2
3–3–2–5–2–3–2/2–3	3–2–3–1–4–4–4/4–1	3–2–1–1–3–2/3–2	5–3–1–2–3–2/5–1

Note. The numbers 1–5 correspond to each of the five possible stimulus elements for the tactile, visual, and auditory modalities (depending on the experimental condition).