Sequential Learning by Touch, Vision, and Audition

Christopher M. Conway (cmc82@cornell.edu)
Morten H. Christiansen (mhc27@cornell.edu)
Department of Psychology
Cornell University
Ithaca, NY 14853, USA

Abstract

We investigated the extent to which touch, vision, and audition are similar in the ways they mediate the processing of statistical regularities within sequential input. While previous research has examined statistical/sequential learning in the visual and auditory domains, few researchers have conducted rigorous comparisons across sensory modalities; in particular, the sense of touch has been virtually ignored in such research. Our data reveal commonalities between the ways in which these three modalities afford the learning of sequential information. However, the data also suggest that in terms of sequential learning, audition is superior to the other two senses. We discuss these findings in terms of whether statistical/sequential learning is likely to consist of a single, unitary mechanism or multiple, modality-constrained ones.

Introduction

The acquisition of statistical/sequential information from the environment appears to be involved in many learning situations, ranging from speech segmentation (Saffran, Newport, & Aslin, 1996), to learning orthographic regularities of written words (Pacton, Perruchet, Fayol, & Cleeremans, 2001) to processing visual scenes (Fiser & Aslin, 2002). However, previous research, focusing exclusively on visual and auditory domains, has failed to investigate whether such learning can occur via touch. Perhaps more importantly, few studies have attempted directly to compare sequential learning as it occurs across the various sensory modalities.

There are important reasons to pursue such avenues of study. First, a common assumption is that statistical/sequential learning is a broad, domain-general ability (e.g., Kirkham, Slemmer, & Johnson, 2002). But in order to adequately assess this hypothesis, systematic experimentation across the modalities is necessary. If differences exist between sequential learning in the various senses, it may reflect the operation of multiple mechanisms, rather than a single process. Second, in regards to the touch modality in particular, prior research has generally focused on low-level perception; discovering that the sense of touch can accommodate complex sequential learning may have important implications for tactile communication systems.

This paper describes three experiments conducted with the aim to assess sequential learning in three sensory modalities: touch, vision, and audition. Experiment 1 provides the first direct evidence for a fairly complex tactile sequential learning capability. Experiment 2 provides a visual analogue of Experiment 1 and suggests commonalities between visual and tactile sequential learning. Finally, Experiment 3 assesses the auditory domain, revealing an auditory advantage for sequential processing. We conclude by discussing these results in relation to basic issues of cognitive and neural organization—namely, to what extent sequential learning consists of a single or multiple mechanisms.

Sequential Learning

We define sequential learning as an ability to encode and represent the order of discrete elements occurring in a sequence (Conway & Christiansen, 2001). Importantly, we consider a crucial aspect of sequential learning to be the acquisition of statistical regularities occurring among sequence elements. Artificial grammar learning (AGL; Reber, 1967) is a widely used paradigm for studying such sequential learning[1]. AGL experiments typically use finite-state grammars to generate the stimuli; in such grammars, a transition from one state to the next produces an element of the sequence. For example, in the grammar of Figure 1, the path begins at the left-most node, labeled *S1*. The next transition can lead to either *S2* or *S3*. Every time a number is encountered in the transition between states, it is added as the next element of the sequence, producing a sequence corresponding to the rules of the grammar. For example, by passing through the nodes *S1*, *S2*, *S2*, *S4*, *S3*, *S5*, the "legal" sequence 4-1-3-5-2 is generated.

During a training phase, participants typically are exposed to a subset of legal sequences—often under the guise of a "memory experiment" or some other such task—with the intent that they will incidentally encode structural aspects of the stimuli. Next, they are tested on whether they can classify novel sequences as

---

[1] In the typical AGL task, the stimulus elements are presented simultaneously (e.g., letter strings)—rather than sequentially (i.e., one element at a time). We consider even the former case to be a sequential learning task because scanning strings of letters generally occurs in a left-to-right, sequential manner. However, our aim here is to create a truly sequential learning environment using temporally-distributed input.

incorporating the same regularities they had observed in the training input. Participants commonly achieve levels of correct classification that are significantly greater than chance. Although there has been disagreement as to what types of information participants use to make correct classification judgments, it is likely that statistical information is an essential piece of the puzzle (e.g., Redington & Chater, 1996). Participants appear to become sensitive to the statistical regularities in the training items—i.e., the frequency with which certain "chunks" of information co-occur—allowing them to generalize their knowledge to novel sequences. It is such statistical sensitivity that we consider to be vital for complex sequential learning tasks.
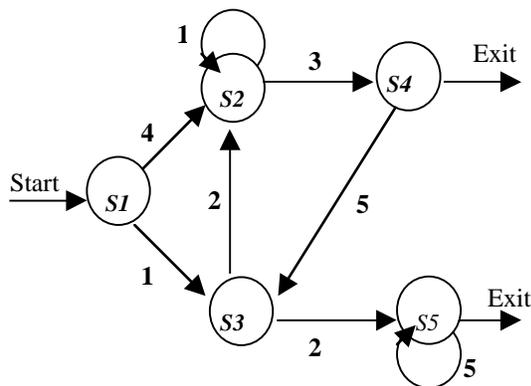

Figure 1: The finite-state grammar used to generate the stimuli for the three experiments.

The standard AGL paradigm has been used extensively to assess visual, as well as auditory learning, suggesting that sequential learning can occur in both modalities. However, two issues remain unexplored: can sequential learning occur in other modalities, such as touch? And, what differences in sequential learning, if any, exist between different sensory modalities?

Experiment 1: Tactile Sequential Learning

The touch sense has been studied extensively in terms of its perceptual and psychophysical attributes (for a review, see Craig & Rollman, 1999), yet only a few studies have hinted that complex sequential learning is possible. For instance, evidence suggests that tactile temporal processing and pattern learning is better than visual, but worse than auditory processing (e.g., Handel & Buffardi, 1969; Manning, Pasquali, & Smith, 1975; Sherrick & Cholewiak, 1986). These studies suggest that touch supports a powerful learning mechanism, which perhaps may be sufficient to allow for successful performance on an AGL task. Experiment 1 attempted to verify this hypothesis.

Method

Participants A total of 20 undergraduates (10 in each condition) from introductory Psychology classes at Southern Illinois University, Carbondale, participated in the experiment. Subjects earned course credit for their participation. The data from an additional five participants were excluded for the following reasons: prior participation in AGL tasks in our laboratory (n=4); did not adequately follow the instructions (n=1).

Apparatus The experiment was conducted using 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, normally used in hand-held paging devices, generated the vibrotactile pulses. Each of these motors was less than 18 mm long and 5 mm wide, making them small enough to be easily attached to the participants' fingers with velcro straps. When activated, the motors produced minor vibrations (rated at 150 Hz) at a magnitude equal to that found in hand-held pagers.

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, allowing precise control over the timing and duration of each vibration stimulus.

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 3 and 6 elements in length. The grammar generates sequences of elements (numbers) with each number being mapped onto a particular finger (1 is the thumb and 5 is the pinky finger). Each tactile stimulus consisted of a sequence of vibration pulses (pulse duration of 250 ms) delivered to the fingers, one finger at a time (250 ms occurring between pulses). For example, the legal sequence 1-2-5-5 corresponds to one vibration pulse delivered to the thumb, then a pulse to the second finger, and lastly two pulses to the fifth finger.

A total of 12 legal sequences, arranged into pairs, were used for training. Six pairs consisted of one training sequence presented twice (matched pairs) whereas the remaining six pairs consisted of two sequences that differed slightly from one another (mismatched pairs). A 2 s pause occurred between the two sequences of each pair.[2]

The test set consisted of ten legal and ten illegal sequences, all of which were novel to the participants. Illegal sequences were produced by beginning each with a legal element, followed by a series of illegal

---

transitions, and ending with a legal element once more. An illegal transition denotes that a particular pair of elements does not occur together during training. 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 do not occur during training). In this manner, the legal and illegal sequences differ from one another in terms of the statistical relationships of adjacent elements.[3]

Procedure Participants were assigned randomly to either a control group or an experimental group. The experimental group participated in both a training and a testing phase, whereas the control group only participated in the testing phase. Before beginning the experiment, participants were assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) to determine their preferred hand. Then, using velcro straps, the experimenter placed a vibration device onto each of the five fingers of the preferred hand.

At the beginning of the training phase, the experimental 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 or not, and indicate their decision by pressing a button marked "YES" or "NO". This match-mismatch paradigm used the twelve training pairs described earlier. It was our intention that this paradigm would encourage participants to pay attention to the stimuli while still allowing incidental learning of the statistical structure to occur.

After the last sequence of each pair, a 1 s pause occurred, followed by a prompt on the screen asking for the participant's response. After the participant made a response, there was a 2 s inter-trial interval before the next pair began.

Each pair was presented six times in random order for a total of 72 exposures, the entire training phase lasting roughly ten minutes. A recording of white noise was played during training to mask the sounds of the vibrators. In addition, the participants' hands were covered by a cardboard box so that they could not visually observe their fingers. These precautions were taken to ensure that tactile information alone, without help from auditory or visual senses, contributed to task performance. As mentioned previously, the experimental group—but not the control group—participated in the training phase.

Before the beginning of the testing phase, the experimental participants were told that the vibration sequences they had just felt had been generated by a computer program that, using a complex set of rules, determined the order of the pulses. They were told that they would now be presented with new vibration sequences. Some of these would be generated by the same program while others would not. It was the participant's task to classify each new sequence accordingly (i.e., whether or not the sequence was generated by the same program) by pressing a button marked either "YES" or "NO." The control participants received the same instructions and task except that there was no reference made to a previous training phase.

The twenty 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 for the training sequences.

Results

The training performance for each experimental participant was assessed by calculating the mean percentage of correctly classified pairs. This calculation revealed that participants, on average, made correct match-mismatch decisions for 74% of the trials.

Results from the testing phase revealed that the control group correctly classified 45% of the test sequences while the experimental group correctly classified 62% of the test sequences. Following Redington and Chater's (1996) suggestions, two analyses were conducted on the 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 performance for each group to chance performance (50%) using single group t-tests.[4]

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, $t(9) = 2.97$, $p < .05$.

The results show that the experimental group significantly outperformed the control group. This suggests that the experimental participants learned aspects of the adjacent element statistics inherent in the training sequences, allowing them to classify novel test sequences appropriately. This is the first empirical evidence of a tactile sequential learning system of such complexity to enable participants to make judgments regarding the legality of artificial grammar-generated sequences.

---

[3]In addition, Gomez and Gerken (1999) matched the legal and illegal sequences in terms of element frequencies and length so that these factors could not influence performance.

[4] Ideally, the control group should perform at chance levels while the experimental group should perform significantly better than both chance and the control group.

Experiment 2: Visual Sequential Learning

Experiment 2 assessed sequential learning in the visual domain. This experiment was identical to Experiment 1 in terms of the general procedure and the timing of the stimuli; however, instead of vibrotactile pulses, the sequences consisted of flashing squares occurring at different spatial locations. The reason for using such stimuli, as opposed to letters, for example, was to provide as close a match as possible to the tactile stimuli used in the first experiment. Importantly, unlike sequences of letters, the vibrotactile sequences consisted of non-linguistic, spatially-distinct elements that were presented one at a time (sequentially). The visual stimuli used for this second experiment shared these same characteristics; therefore, the resulting data should provide a meaningful basis for comparison with the first experiment. Like Experiment 1, there was an experimental group, undergoing training and testing phases, and a control group, undergoing the testing phase only.

Only a handful of statistical learning studies have used non-linguistic visual stimuli in a truly sequential manner (e.g., Fiser & Aslin, 2002; Kirkham et al, 2002). The data suggest that such a presentation does not hamper sequential learning by vision. However, other studies (e.g., Handel & Buffardi, 1969) indicate that for certain temporal processing and pattern learning tasks, vision may be inferior to touch. This experiment aimed to investigate whether such differences would be observed.

Method

Participants An additional 20 undergraduates (10 in each condition) were recruited from introductory Psychology classes at Cornell University. Subjects received extra credit for their participation. The data from three additional participants were excluded for the following reasons: did not adequately follow the instructions (n=2); equipment malfunction (n=1).

Apparatus The apparatus was the same as Experiment 1, except for the exclusion of the vibration devices.

Materials The sequences were identical to those of Experiment 1 except that instead of vibrotactile pulses, they were composed of flashing black squares displayed on the computer monitor (1 was the leftmost location and 5 was the rightmost). Each flashing square appeared for 250 ms and was separated by 250 ms. Thus, 1-2-5-5 represents a sequence consisting of a flash appearing in the first location, then in the second location, followed by two flashes in the fifth location.

Procedure The procedure was the same as that of Experiment 1, the only differences relating to the nature of the stimuli presentations, as described above. The timing of the stimuli were identical to those of Experiment 1.

Results

The same statistical analyses as used in Experiment 1 were performed. During the training phase, the experimental group participants made correct match-mismatch decisions on 86% of the trials. A comparison of means across the two experiments revealed a significantly higher training performance in Experiment 2, $F(1, 18) = 14.21, p < .01$.

Results for the testing phase revealed that the control group correctly classified 47% of the test sequences while 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, $t(9) = 3.03, p < .05$.

The results indicate that the experimental group significantly outperformed the control group. In addition, overall experimental and control group performance at test was very similar to that observed in Experiment 1, suggesting commonalities between tactile and visual sequential learning.

Experiment 3: Auditory Sequential Learning

Experiment 3 assessed sequential learning in the auditory domain. This experiment was identical to Experiments 1 and 2 except that it used sequences of auditory tones. Like the previous experiments, Experiment 3 had an experimental group, undergoing training and testing phases, and a control group, undergoing the testing phase only. Although previous research has found similar statistical learning performance in vision and audition (Fiser & Aslin, 2002), other data suggest that audition excels at sequential processing tasks (Handel & Buffardi, 1969; Sherrick & Cholewiak, 1986); therefore, we might expect to see a difference in auditory compared to visual and tactile sequential learning.

Method

Participants An additional 20 undergraduates (10 in each condition) were recruited from introductory Psychology classes at Cornell University.

Apparatus The apparatus was the same as Experiment 2. The auditory tones were generated using the SoundEdit 16 version 2 software for the Macintosh.

Materials The sequences were identical to those used in the previous experiments except that instead of vibrotactile pulses or flashing black squares, they consisted of musical tones beginning at middle C (1 = C, 2 = D flat, 3 = F, 4 = G flat, and 5 = B).[5] Each tone lasted 250 ms and was separated by 250 ms. Thus, the sequence 1-2-5-5 consists of a C, then a D flat, and lastly two B's.

Procedure The overall procedure was the same as that of the previous experiments.

Results

During the training phase, the experimental group participants made correct match-mismatch decisions on 96% of the trials. This training performance was significantly higher than that of Experiment 2, $F(1, 18) = 10.20, p < .01$.

Results for the testing phase revealed that the control group correctly classified 44% of the test sequences while 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, $t(9) = 7.45, p < .001$.

Like the previous experiments, the data indicate that the experimental group significantly outperformed the control group; hence, participants appeared to learn aspects of the statistical structure of the input. In fact, the experimental group test performance appears to be substantially greater compared to those of Experiments 1 and 2 (75% vs. 62% and 63%).

General Discussion

Assessing first the training results, we found that performance was significantly different across all three experiments (audition = 96%; vision = 86%; touch = 74%). Because the training task essentially involves remembering and comparing sequences within pairs, the results may elucidate possible differences between the three modalities in representing and maintaining sequential information (Penney, 1989). It is also possible that these results are due to factors such as differential discriminability or perceptibility of sequence elements in different sensory domains.

The testing results for all three experiments are summarized in Figure 2. All three experiments are similar in that the experimental group test performances were significantly different from both chance and their respective control groups. From these results, it appears that participants learned aspects of the adjacent element statistical structure inherent in the training input, allowing them to classify novel stimuli. In this manner, tactile, visual, and auditory sequential learning display commonalities. It is especially interesting to note that sequential learning is not limited to the visual and auditory modalities, but extends to touch as well.
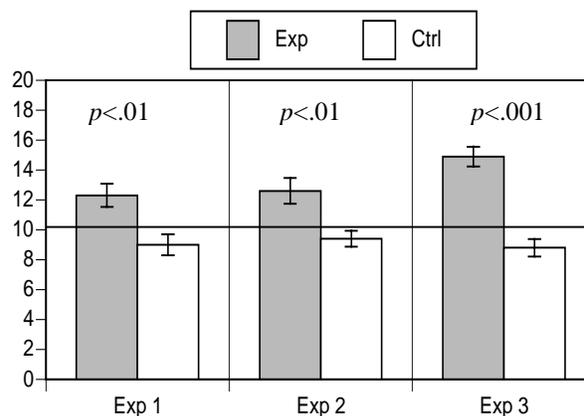

Figure 2: Summary of test results (# of correct responses out of 20).

Despite this overall similarity across modalities, it is also apparent that the Experiment 3 (auditory) results are somewhat different from the other two experiments. Specifically, the auditory experimental group performed better at test as compared to the tactile and visual experimental groups (75% vs. 62% and 63%). This difference is in fact significant [$F (1, 54) = 6.03, p < .05$].[6] Thus, it appears that in this task, auditory sequential learning was more successful than both tactile and visual learning. While previous research has suggested that audition excels at relatively low-level temporal processing tasks (Mahar et al., 1994; Sherrick & Cholewiak, 1986), our results appear to be the first evidence that such an advantage extends to complex temporal processing, namely statistical/sequential learning. This auditory advantage perhaps is related to the finding that adults process tone sequences by representing relative, as opposed to absolute, pitch (Saffran & Griepentrog, 2001); such a strategy may allow for more efficient encoding of adjacent element statistics.

---

[5] This particular set of notes was used because it avoids familiar melodies.

[6] This was computed by contrasting the means of the experimental and control groups, as illustrated by the equation: E3-C3 = .5(E1-C1)+.5(E2-C2), where E and C refer to experimental and control group means, respectively.

Conclusion

It has been argued that statistical learning is subserved by a single, domain-general mechanism (e.g., Kirkham et al., 2002). Although a single-mechanism view may be theoretically attractive, our results point toward another possibility: that sequential learning may involve multiple, modality-constrained processes. This idea is supported by a recent multivariate meta-analysis of 35 PET experiments (Lloyd, 2000), which suggested that computations in the different "sensory streams" (i.e., representations of tactile, visual, and auditory information) rely on entirely different cortical areas altogether, at all levels of processing. Additionally, neuroimaging evidence specifically related to sequential learning is consistent with a multiple-mechanism view (see Clegg, DiGirolamo, & Keele, 1998). Thus, we propose that sequential learning is best understood as a functional "suite", composed of multiple, modality-constrained mechanisms. Each mechanism is instantiated in largely non-overlapping brain areas but some degree of interaction is likely to occur between them. We further suggest that each modality-constrained mechanism shares similar computational properties with one another, including the ability to extract adjacent element statistics from incidental exposure to input. However, because each learning mechanism is largely tied to specific sensory areas, each is constrained by the global properties of that sensory system. These properties presumably relate to the types of information that each sensory modality is specialized to process, such as temporal, spatiotemporal, or spatial configurations (Mahar et al., 1994). Our experimental data illustrate one example of such specialization: the auditory system encoded statistical information of temporal input more effectively than did the other senses. Important targets for future research include further substantiating this multiple mechanism view of sequential learning and to discover how such modality-constrained systems might interact with each other, as well as how each relates to human cognition in general. We anticipate that such future research, especially that involving neurophysiological experimentation, will further elucidate the nature of sequential learning by touch, vision, and audition.

References

Clegg, B.A., DiGirolamo, G.J., & Keele, S. (1998). Sequence learning. *Trends in Cognitive Sciences, 2*, 275-281.

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.

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.

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.

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.

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.

Lloyd, D. (2000). Terra cognita: From functional neuroimaging to the map of the mind. *Brain & Mind, 1*, 93-116.

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

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.

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.

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

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

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

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*). New York: Wiley & Sons.