Received: 14 November 2017 | Accepted: 12 January 2018
DOI: 10.1002/dev.21614

RESEARCH ARTICLE

WILEY Developmental Psychobiology

Learning of hierarchical serial patterns emerges in infancy

David J. Lewkowicz[1] | Mark A. Schmuckler[2] | Diane M.J. Mangalindan[2]

[1] Department of Communication Sciences and Disorders, Northeastern University, Boston, Massachusetts

[2] Department of Psychology, University of Toronto Scarborough, Scarborough, Ontario, Canada

Correspondence
David J. Lewkowicz, Department of Communication Sciences and Disorders, Northeastern University, Boston, 02115 MA.
Email: d.lewkowicz@northeastern.edu

Funding information
National Science Foundation, Grant number: BCS-0751888; Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant number: R01HD057116; Discovery Grant, Natural Sciences and Engineering Research Council of Canada

Abstract

Recursive, hierarchically organized serial patterns provide the underlying structure in many cognitive and motor domains including speech, language, music, social interaction, and motor action. We investigated whether learning of hierarchical patterns emerges in infancy by habituating 204 infants to different hierarchical serial patterns and then testing for discrimination and generalization of such patterns. Results indicated that 8- to 10-month-old and 12- to 14-month-old infants exhibited sensitivity to the difference between hierarchical and non-hierarchical structure but that 4- to 6-month-old infants did not. These findings demonstrate that the ability to perceive, learn, and generalize recursive, hierarchical, pattern rules emerges in infancy and add to growing evidence that general-purpose pattern learning mechanisms emerge during the first year of life.

KEYWORDS
attention, human infants, learning, pattern learning, perception

1 | INTRODUCTION

Humans and animals exhibit a remarkable ability to perceive, learn, produce, and remember hierarchically organized serial patterns (Fountain & Rowan, 1995; Garlick, Fountain, & Blaisdell, 2017; Greenfield, 1991; Lashley, 1951; Martin, 1972; Restle, 1970; Restle & Burnside, 1972). These types of patterns are characterized by the rule-governed, recursive embedding of lower-level units (chunks) into higher-level units. In general, hierarchical patterns have unlimited depth because recursion permits the generation of an infinite number of combinations from a finite number of elements. As a result, such patterns can generate highly complex structures that are efficient, highly predictable, and easier to learn and remember than unpatterned collections of the same set of elements (Restle & Burnside, 1972). It is probably for this reason that recursion and hierarchical patterning play such a fundamental role in human speech, language, vision, music, event perception, and motor action (Greenfield, 1991; Hauser, Chomsky, & Fitch, 2002; Jackendoff & Pinker, 2005; Zacks & Tversky, 2001).

Given the power of recursion and hierarchical patterning for representing information, it would be highly adaptive if the ability to perceive and learn hierarchical serial patterns emerged early in development. Infancy is a time of rapid growth in many domains of cognitive functioning and, thus, it would not be surprising if this ability emerged at this time. Unfortunately, with the exception of one study (Werchan, Collins, Frank, & Amso, 2015), there have been no investigations of this question in infancy. The vast majority of studies to date on infant perception and learning of patterned information have focused on relatively low-level aspects of pattern learning that is required but not sufficient for the detection of recursive, hierarchical, serial patterns. For example, studies have found that infants can perceive the rhythm/prosody of their native language (Nazzi, Bertoncini, & Mehler, 1998) as well as the rhythmic structure of non-speech auditory (Chang & Trehub, 1977; Demany, McKenzie, & Vurpillot, 1977) and audiovisual patterns (Lewkowicz & Marcovitch, 2006; Pickens & Bahrick, 1997). Moreover, studies have found that many months before frank language production and perception begin to emerge, infants begin exhibiting the ability to perceive and learn the statistics (i.e., conditional probability relations) that link both adjacent and non-adjacent items in various types of sequences. For example, it has been found that infants can learn adjacent and non-adjacent statistics regardless of whether the statistics specify sequences consisting of nonsense syllables (Gómez, 2002; Gómez & Maye, 2005; Saffran, Aslin, & Newport, 1996), musical tones (Saffran, Johnson, Aslin, & Newport, 1999), abstract shapes (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Kirkham, Slemmer, Richardson, & Johnson, 2007; Marcovitch & Lewkowicz, 2009), or abstract shapes and sounds (Lewkowicz & Berent, 2009). Finally, it has been found that infants can perceive and learn ordinal sequential relations specified either by words (Mandel, Nelson, & Jusczyk, 1996) or abstract objects and their sounds (Lewkowicz, 2013) and that they can also learn simple reduplication rules such as AAB or ABA (Gerken, 2006; Gómez & Gerken, 1999; Marcus, Vijayan, Rao, & Vishton, 1999; Saffran, Pollak, Seibel, & Shkolnik, 2007).

There is little doubt that the above-referenced pattern-learning skills are necessary for the learning of hierarchical serial patterns. Unfortunately, however, these skills are not sufficient for learning hierarchical structure per se. As indicated earlier, patterns are characterized by the rule-governed, recursive embedding of lower-level units (chunks) into higher-level units. This is made possible by the fact that serial patterns are defined by much more than the simple, linear temporal relationship of two sequence elements. Serial patterns are defined by a formal hierarchical structure that specifies a systematic relationship of individual pattern elements with lower-order rules and a structure that specifies the relationship of the lower-order rules according to a set of higher-order rules. This sort of hierarchical patterning is characteristic of natural languages where phrases are recursively embedded within phrases in a rule-bound manner, creating long-distance relationships. Interestingly, evidence indicates that infants and children have the capacity to perceive and learn patterns defined by hierarchical rules. This evidence comes from studies that are based on work of Badre, D'Esposito, and colleagues (Badre, 2008; Badre & D'Esposito, 2007; Badre, Kayser, & D'Esposito, 2010; Kayser & D'Esposito, 2013). This work indicates that children have the ability to use hierarchical rule sets to guide response categorizations (Amso, Haas, McShane, & Badre, 2014; Unger, Ackerman, Chatham, Amso, & Badre, 2016; Werchan et al., 2015). One of these studies is particularly relevant in the current context because Werchan et al. (2015) investigated whether 8-month-old infants can learn hierarchical rule sets. Findings indicated that infants can, indeed, learn such rules, that they can use them to guide their oculomotor responses, that they can associate specific rules with specific objects, and that they can generalize these rules to novel contexts.

The sort of learning skill uncovered by Werchan et al. (2015) demonstrates for the first time that infants as young as 8 months of age are capable of higher-level, associative learning based on hierarchically structured rules. It should be noted, however, that it is not clear whether the sort of learning examined in the Werchan et al. (2015) study extends to spatiotemporally extended events because spatio-temporal information was not manipulated in this study. In essence, Werchan et al. (2015) demonstrated that 8-month-old infants can learn specific rules governed by higher-order contexts for associating arbitrary object properties (i.e., color and shape) with location as well as associating different objects with different sets of words depending on their face-voice context. Specifically, these investigators found that 8-month-old infants could attach different sets of labels to identical sets of objects depending on whether those objects were associated with one person's face and voice or another person's face and voice. For example, the infants learned that a duck was associated with the word "tiv" and that a rattle was associated with the word "fep" when the two objects were presented in the context of one person's face and voice but that these two objects were associated with the words "mip" and "dax," respectively, when they were presented in the context of another person's face and voice. Finally, and most importantly, Werchan et al. (2015) showed that their infants generalized their learning to novel higher-order contexts (i.e., novel faces and voices), indicating that 8-month-old infants can learn latent, 2-level, hierarchical rule sets.

The question of whether infants might be able to learn spatiotemporally based hierarchical patterns is important because infants live in a spatiotemporally extended world that is full of discrete events. For example, when infants hear someone speaking or playing music, they can hear temporally structured serial patterns. When they can also see someone speaking or playing music, they can now see and hear spatiotemporally structured serial patterns. Extraction of the meanings inherent in these sorts of events often depends on the ability to extract the rule-governed spatiotemporal distribution of the elements that constitute particular types of events. This is, in essence, what Lashley (1951) referred to as the problem of serial order in his classic paper by the same title. Since Lashley's statement of the problem, many others have investigated this problem under the rubric of serial pattern learning (e.g., Greeno & Simon, 1974; Jones, 1976, 1981; Kovotsky & Simon, 1973; Martin, 1972; Restle, 1970, 1972; Restle & Brown, 1970; Restle & Burnside, 1972; Simon, 1972; Simon & Kotovsky, 1963; Vitz & Todd, 1969).

One of Lashley's (1951) key observations was that the hierarchical organization inherent in serial patterns only emerges if the information that specifies the precise occurrence of the elements that constitute specific events is integrated. Currently, it is not known whether infants can integrate the spatiotemporal information inherent in hierarchically organized serial patterns because the only study to date to have examined hierarchical rule learning (Werchan et al., 2015) did not test infants' ability to learn spatiotemporally determined rules. Therefore, the goal of the current study was to investigate whether infants can learn and generalize spatiotemporally generated recursive hierarchical serial patterns and, if so, when this ability might first emerge.

Predicting when this ability might first emerge is somewhat difficult in the absence of relevant evidence. Based on Werchan et al. (2015), one might expect this ability to emerge around 8 months of age. It might be argued, however, that the perception of hierarchical structure inherent in serial patterns is more difficult because the integration of information over space and time requires greater cognitive resources than does the learning of associative rules. If this is true then the ability to learn hierarchical serial patterns might not emerge until later in development. Consequently, the most sensible approach is to test infants across a wide-enough age range to allow for the possibility that this ability may not emerge until after 8 months of age but, at the same time, to allow for the possibility that it might emerge even earlier. Hence, in Experiment 1, we examined perception and learning of hierarchical serial patterns in 8- to 10-month-old and 12- to 14-month-old infants while in Experiment 2 we investigated this ability in infants as young as 4–6 months of age in case this ability actually emerges earlier.

We had two goals in mind. One was to investigate whether infants could learn specific hierarchical serial patterns. The other was to investigate whether infants could also generalize their learning of such patterns to novel hierarchical patterns by abstracting the general concept of hierarchical organization. To accomplish these two goals, we used a habituation procedure that is specifically designed to encourage infants to learn the specific properties of a given stimulus as well as a set of rules that specify a specific category (Cohen, 2009; Cohen & Strauss, 1979). This so-called "multiple habituation" procedure involves habituating infants to multiple exemplars of a particular event. In the present case, we habituated infants to three distinct hierarchical serial patterns by presenting each during separate habituation trials. This way, we provided infants with the opportunity to learn about each specific pattern as well as the general concept of hierarchical serial patterning. Then, once infants reached the habituation criterion, we administered two sets of test trials. In one set, we tested whether the infants learned the specific patterns by examining their ability to discriminate between a hierarchical and non-hierarchical version of one of the familiar patterns. In the other set, we tested whether the infants also learned the general concept of hierarchical serial patterning by examining their ability to discriminate between a novel hierarchical serial pattern and its non-hierarchical version.

The three different hierarchical serial patterns presented in the current study consisted of the actions of a series of identical objects and sounds. To control the specific actions depicted in these patterns and to create serial hierarchical structure, we employed a method that has been used widely in work on serial pattern learning (e.g., Greeno & Simon, 1974; Jones, 1976, 1981; Kovotsky & Simon, 1973; Restle, 1970, 1972; Restle & Brown, 1970; Restle & Burnside, 1972). This method involved generating sequences of stimuli according to the rules underlying binary hierarchical tree structures which, in turn, involves the use of a set of compound algebraic functions that define serial transitions. In our case, these transitions were repetition, next, and complement. Typically, these types of algebraic functions produce hierarchically organized symmetrical-tree patterns that consist of low-level rule-generated units that are incorporated into increasingly higher-level rule-generated units (Restle, 1970). Classic research with these types of patterns has shown that human adults learn them as abstract rule-governed entities (Restle, 1972). More recent work with rats and pigeons has demonstrated that animals also can learn such abstract rule-governed entities (Fountain, 2008; Fountain & Doyle, 2011; Fountain & Rowan, 1995; Garlick et al., 2017; Muller & Fountain, 2016).

Our explicit aim was to test the possibility that learning of hierarchical serial patterns in infancy is a domain-general skill rather than one tied to a specific domain. Therefore, we presented patterns that were instantiated by the actions of arbitrary objects and arbitrary (i.e., non-speech) sounds. During the habituation phase, we repeatedly presented three distinct hierarchical serial patterns until infants reached a pre-defined habituation criterion. Once they reached the criterion, we administered a set of Familiar test trials and a set of Generalization test trials (links to sample videos of some of the patterns presented in this study and their description can be can be found in Supporting Information). The Familiar set was designed to determine whether infants successfully learned a specific hierarchical pattern and whether they could discriminate it from the identical but non-hierarchical version of the same pattern instantiated by the identical stimuli. Thus, during the Familiar test set, we contrasted responsiveness to one of the previous three hierarchical serial patterns versus their yoked non-hierarchical versions. Yoking meant that infants saw the same stimuli and heard the same sounds except that this time the serial pattern was not hierarchical. The Generalization set was designed to determine whether infants abstracted the concept of hierarchical organization and, thus, whether they could discriminate between a novel hierarchical pattern—one that they had not experienced before and that was instantiated with novel visual stimuli and sounds—and a novel and yoked non-hierarchical version of this same novel pattern. We expected the Familiar test trials to yield a significant response recovery if the infants successfully learned one of the specific hierarchical serial patterns. Similarly, we expected the Generalization test trials to yield a significant response recovery if the infants successfully abstracted the concept of hierarchical patterning during the habituation trials and then generalized this concept to the discrimination of a novel hierarchical serial pattern from its non-hierarchical version.

2 | EXPERIMENT 1

2.1 | Method

2.1.1 | Participants

We tested seventy-two 8- to 10-month-old infants (M = 8.81 months; SD = 1.029; 30 girls) and eighty-nine 12- to 14-month-old infants (M = 13.03 months, SD = 1.11; 48 girls). We also tested 48 additional infants but did not include their data due to their being the wrong age (n = 1), health concerns (n = 3), refusal to complete the experiment (n = 2), fussiness (n = 30), inattentiveness (n = 2), experimental error (n = 7), equipment failure (n = 1), parental interference (n = 1), and difficulty in coding looking behavior (n = 1). We tested infants in two different laboratories, with most of the infants, including those tested in Experiment 2, tested at Florida Atlantic University (FAU; n = 177) and the rest at the University of Toronto Scarborough (UTSC).

2.1.2 | Apparatus and stimuli

Infants were tested either in a sound-attenuated booth (FAU) or in a quiet experimental room (UTSC). Most were seated in an infant seat, although some were seated on the parent's lap. When parents held their infants they were not aware of the hypothesis under test, were asked to sit still and refrain from any interactions with their infant, and either wore headphones and listened to white noise (FAU) or were seated facing a blank wall (UTSC). The infants were seated between 50 and 60 cm from a 17-inch computer monitor. The audio portion of the stimulus events was presented through a pair of speakers located on each side of the stimulus presentation monitor. A video camera, focused on an infant's face, was positioned either on top of (FAU) or adjacent to (UTSC) the monitor and recorded infant looking behavior. An experimenter was seated outside the testing booth (FAU) or in an adjacent area (UTSC) and controlled a computer that presented all stimulus events. The experimenter, who was blind with respect to stimulus presentation, observed the infants' looking behavior on a monitor connected to the camera focused on the infant's face and presented stimulus events contingent on the infants' looking at the stimulus monitor.

All stimulus events were presented as multimedia movies. One of these movies was an attention-getter which consisted of a continuously and silently expanding and contracting green disk. Its purpose was to attract the infants' attention to the stimulus-presentation monitor. A second movie was a clip of a Winnie-the-Pooh cartoon that could be seen and heard. Its purpose was to measure initial and terminal attention to rule out fatigue effects. The remaining movies consisted of horizontal arrays of six eggs and associated sounds. These movies were used to produce different serial patterns by activating the eggs in temporally distinct ways. Each trial began with the appearance of the array of the six eggs followed by their openings and closings accompanied by a sound. Each egg opening/closing lasted 1 s and consisted of the following actions: the egg increased slightly in size, the top half of the egg shell moved up, a chick popped up from the bottom shell and made a concurrent sound, the chick disappeared, the top shell came back down, and the egg resumed its original size. Each trial involved the serial activation of different eggs until an 8-element temporal pattern was generated (this meant that some eggs were activated more than once in a given pattern). Table 1 shows the various types of serial patterns generated for the different trials. As can be seen, there were two sets of patterns: one set consisted of the eggs opening left-to-right and the other set consisted of the eggs opening right-to-left. Each set, in turn, consisted of three Familiar hierarchical patterns and three non-hierarchical versions of those patterns as well as a Generalization hierarchical pattern and a non-hierarchical version of it.

The specific temporal sequence of egg openings for each type of trial depicted in Table 1 was determined by a set of nested transformation rules that resulted in a nested pattern structure. The rules used to generate the different patterns also can be seen in Table 1 (their actual implementation is explicated in greater detail below). To prevent infants from discriminating the hierarchical patterns from their non-hierarchical version on the basis of spatial cues—rather than temporal organizational cues—the non-hierarchical versions of each corresponding hierarchical pattern were generated by yoking the egg openings in these patterns to the egg openings in the hierarchical patterns. This meant that the identical eggs opened, except that this time they did so in a manner that was not hierarchically organized. In addition, to avoid primacy and/or recency effects and, thus, to prevent the infants from relying on the first and/or last pattern elements for discrimination between the hierarchical and non-hierarchical patterns, we ensured that the first and eighth egg openings of the non-hierarchical patterns were the same as in the corresponding hierarchical patterns.

The sound that accompanied the Familiar hierarchical and the corresponding non-hierarchical patterns was a short peep while the sound that accompanied the Generalization hierarchical and the corresponding non-hierarchical patterns was a quack. Figures 1a and 1b provide an example of the visual stimuli used to instantiate one of the Familiar hierarchical and non-hierarchical patterns, a schematic

TABLE 1 The hierarchical and non-hierarchical rules and the patterns generated by these rules

	Hierarchical rules	Hierarchical pattern	Non-hierarchical rules	Non-hierarchical pattern
Left-to-right egg opening				
Familiar	$N^1(C(N^2(1)))$	13642453	$(N^3(1))(R)(N^{-1})(N^{-1})(N^3)(N^1)(N^{-3})$	14432563
Familiar	$C(R(N^1(1)))$	12126565	$(N^5(1))(N^{-4})(R)(N^3)(N^{-4})(N^5)(N^{-1})$	16225165
Familiar	$N^2(N^2(N^{-1}(2)))$	21434365	$(N^1(2))(N^1)(N^{-1})(N^3)(N^{-2})(N^{-3})(N^4)$	23436415
Generalization	$C(N^2(R(2)))$	22445533	$(N^2(2))(N^{-1})(N^{-1})(N^2)(N^{1)}(R)(N^{-2})$	24324553
Right-to-left egg opening				
Familiar	$R(N^{-1}(C(6)))$	61526152	$(N^{-4}(6))(N^{-1})(N^4)(R)(N^1)(N^{-5})(N^1)$	62155612
Familiar	$C(N^{-1}(N^{-1}(6)))$	65541223	$(N^{-5}(6))(N^1)(N^2)(N^1)(N^{-3})(N^3)(N^{-2})$	61245253
Familiar	$C(N^1(N^{-2}(4)))$	42533524	$(N^{-1}(4))(N^{-1})(N^1)(N^2)(R)(N^{-3})(N^2)$	43235524
Generalization	$C(C(N^2(4)))$	46313146	$(N^{-1}(4))(N^{-2})(R)(N^2)(N^3)(N^{-2})(N^2)$	43113646

The top panel shows the set of patterns where the eggs opened and sounded from left-to-right while the bottom panel shows the set of patterns where the eggs opened and sounded from right-to-left. The numbers in the Pattern columns refer to which of the six eggs was opened, in turn, based on an 8-element sequence of events and should be read from left to right. Familiar refers to patterns presented during the habituation phase and during the Familiar set of test trials and Generalization refers to patterns presented during the Generalization test trials. For specific details of how the rules were implemented to generate the patterns specified in the table please see the Section 2.1.3.

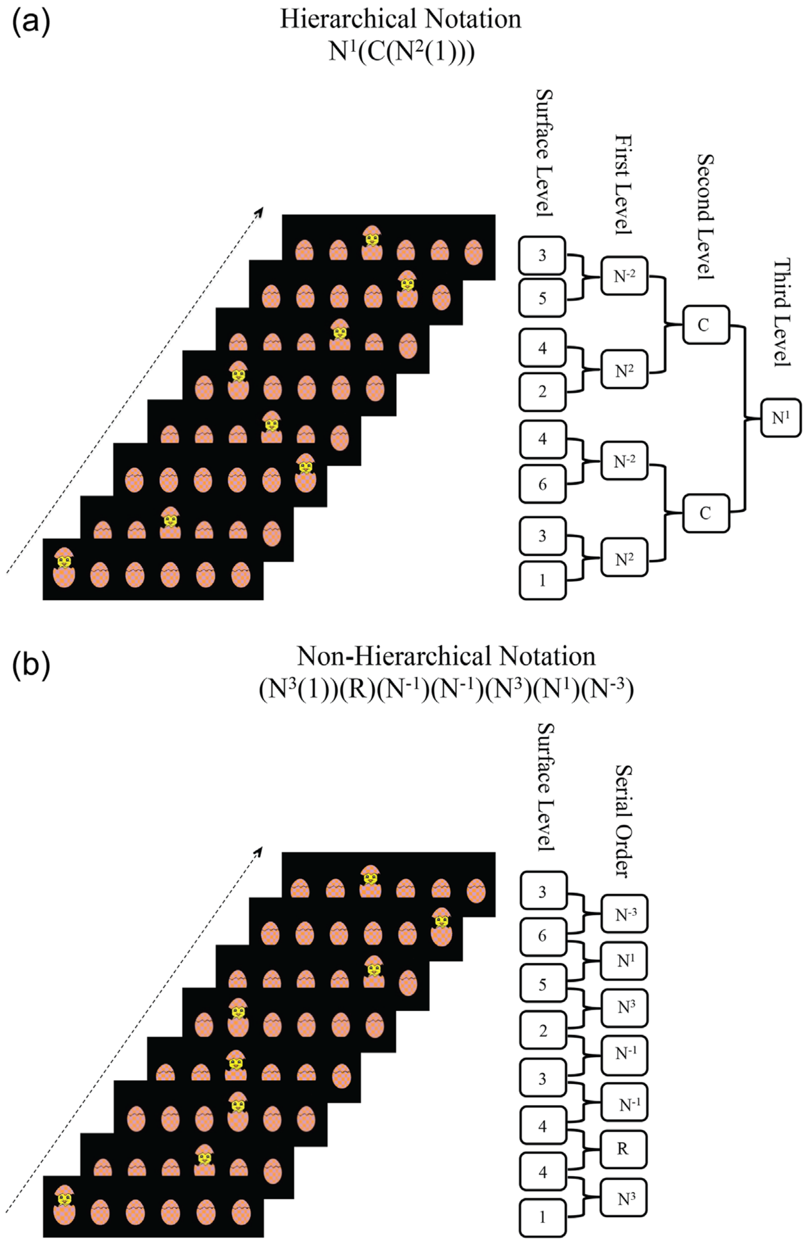

FIGURE 1 Schematic representation of one of the hierarchical patterns (a) and its yoked, non-hierarchical version (b). The left side of panels (a) and (b) depicts which specific eggs were opened across time during a single iteration of the first left-to-right Familiar hierarchical and non-hierarchical patterns depicted in Table 1. On the top of each respective pattern can be seen the formula that specified the transition rules used to generate the patterns and on the right of each respective pattern can be seen the 3-level hierarchical and the 1-level non-hierarchical patterns that each respective rule generated

of their serial presentation, and the rules used to generate each pattern. Figure 2 shows the differences in the eggs and chicks used to generate the Familiar and the Generalization patterns, respectively. As Figure 2 shows, the Generalization patterns differed from the Familiar patterns in that they were instantiated by yellow, rather than orange/purple eggs, and deep orange, rather than yellow, chicks. By instantiating the Generalization patterns with completely novel stimuli, we were able to determine whether infants were able to generalize their learning beyond the specific stimuli that instantiated the familiar patterns.

2.1.3 | Implementation of pattern rules

The formulas associated with the rules used for generating the different patterns can be seen in Table 1. The formulas specify the transition rules of pattern elements including repetition (R), next (N), and complement (C; mirror image). Application of these rules respects the customary mathematical order of operations. The superscripts indicate the step increase made to the starting number between 1 and 6. Positive superscripts indicate addition; negative superscripts indicate subtraction. The starting point is indicated by the number within the parenthesis next to the innermost or first transition rule.

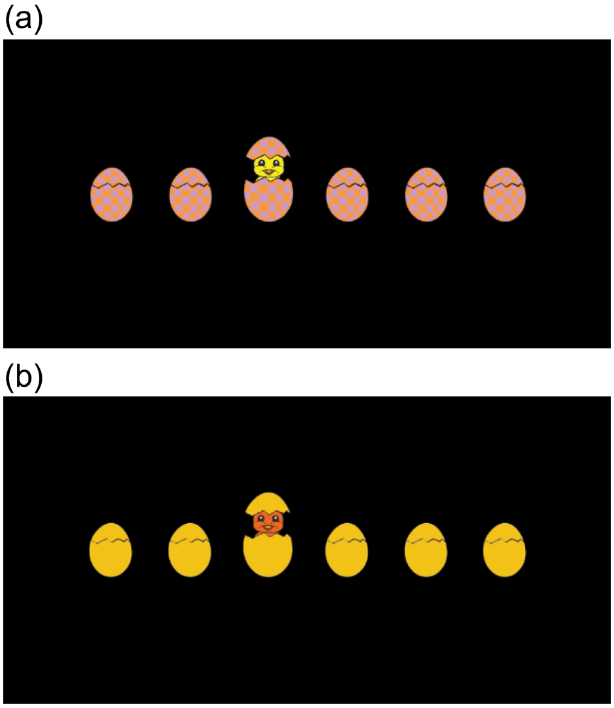

FIGURE 2 Comparison of the eggs and chicks used for generating the Familiar and Generalization patterns. As can be seen, the eggs used for the Familiar patterns (a) were colored with orange and purple alternating squares whereas the eggs used for the Generalization patterns (b) had a uniform yellow color. Also, the chicks used for the Familiar patterns were yellow whereas those used for the Generalization patterns were deep orange. Finally, the sounds presented together with the Familiar patterns were peeps whereas the sounds presented together with the Generalization patterns were quacks

To illustrate the implementation of the rules, consider the 8-element pattern specified by the first Familiar hierarchical pattern in Table 1 and further illustrated in Figure 1: $N^1(C(N^2(1)))$. First, assume a starting number of 1. Then, based on order of operations, the N^2 rule within the innermost brackets is processed. The rule requires the addition of 2 to the starting number. This creates a 2-element chunk consisting of 13. Because this chunk is nested with the next innermost rule, C, this rule operates on the previous chunk as a whole and, thus produces the complement (i.e., mirror image) of 13 in a sequence comprised of six elements. This means, start with a 6 and then subtract 2. This process is indicated in Figure 1 by the application of N^{-2}, which is the complement of the N^2 rule as given in the original formula. This yields 64 and when combined with the previous chunk, results in the 4-element chunk of 1364. This entire chunk can be seen as the bottom half of the schematized version of the pattern in Figure 1. Finally, this bottom chunk is then nested with the final rule, N^1, resulting in a repeat of the rules used to generate the bottom chunk in the Figure except that now we apply the last rule (i.e., N^1). Accordingly, we add a 1 to the first element of the bottom chunk in Figure 1, and then apply the previous two rules, namely the N^2 rule (generating the sequence 24) and then the C rule (generating the sequence 53), to produce the sequence 2453. This chunk is then concatenated with the previous 4-element chunk to produce the complete sequence 13642453. This particular pattern specifies which of the six horizontally distributed elements should be activated and the order in which each should be activated, starting with the left-most element first.

To illustrate the implementation of the rules for creating a non-hierarchical pattern, consider the first such rule in Table 1: $(N^3(1))(R)(N^{-1})(N^{-1})(N^3)(N^1)(N^{-3})$. This type of pattern contains a serial structure but no nested relations. As a result, the pattern is processed strictly from left to right. Again, assume a starting point of 1. Then, based on the rule, N^3, add 3 to the initial element, producing the 2 element chunk 14. Because it is not nested (i.e., hierarchical), the R operates on only the final element of this chunk, repeating this element, thus resulting in the 3-element chunk 144. The N^{-1} operator decreases the final element by 1, producing the 4 element chunk 1443, followed by another N^{-1} operator which produces the 5 element chunk 14432. The N^3 operator adds 3 to the final element, producing the 6 element chunk 144325, then N^1 operator adds 1 to the final element producing the 7 element chunk 1443256, and the N^{-3} operator subtracts 3 from the final element producing the 8 element 14432563.

2.1.4 | Procedure

We used an infant-controlled habituation procedure that enabled infants to control the start, the duration, and the end of each trial with their looking behavior. Specifically, whenever infants looked at the stimulus-presentation monitor, a particular trial began and continued until the infants looked away from the monitor for more than 1 s or until 53 s elapsed. At this point, the attention-getter appeared on the screen and the next trial began when the infants looked back at it. An experimenter, who was blind to the stimuli being presented, recorded the infants' looking behavior by watching them on a monitor that transmitted a view of their face from a camera placed above the stimulus–presentation monitor.

The habituation phase was the first part of the experiment. It began with a pretest trial during which the Winnie-the-Pooh movie was presented to obtain a baseline measure of the infant's initial level of attention. Once this trial ended, the habituation trials started and continued until the habituation criterion was met. The criterion required that an infant's total amount of looking during the last four habituation trials decline to 50% of the total amount of looking during the first four habituation trials. To give infants the opportunity to not only learn specific hierarchical serial patterns but to also learn the abstract concept of hierarchical patterning, we used the multiple habituation procedure and, thus, habituated them to three different hierarchical serial patterns. One group at each age was habituated to the three hierarchical patterns instantiated by the eggs opening from left to right (i.e., the three Familiar hierarchical patterns depicted in the top part of Table 1). The other group was habituated to the three hierarchical patterns instantiated by the eggs opening from right to left (i.e., the three Familiar hierarchical patterns depicted in the bottom part of Table 1). Each pattern was presented separately during a given habituation trial and was repeated continuously until the infant either looked away for 1 s or until 53 s elapsed. For each infant, the three different Familiar hierarchical patterns were presented in sets of three, with the order of the individual patterns within each set based on a Latin square design that generated a total of nine trials. If a given infant did not reach the habituation criterion by the end of the first set of nine trials, the entire set of nine trials was begun anew. This continued until the infant reached the habituation criterion.

The test phase was the second part of the experiment. It began as soon as infants reached the habituation criterion and as soon as they looked back at the monitor following a look-away response. It consisted of two sets of two test trials each. Once again, a specific pattern was presented repeatedly during each of these trials until the infant met the look-away criterion. The first set was the Familiar set of test trials. In this set, responsiveness to one of the three Familiar hierarchical patterns presented during the habituation phase was contrasted to responsiveness to its corresponding non-hierarchical version. This enabled us to determine whether infants detected a disruption of the hierarchical patterning of one of the Familiar patterns. The second set of test trials was the Generalization set. During this set of test trials, responsiveness to a novel hierarchical pattern not experienced before was compared with responsiveness to the corresponding and also novel non-hierarchical pattern. This enabled us to determine whether infants learned the abstract concept of hierarchical structure and, thus, whether they encoded the rules relating to such patterning.

The test phase began with the presentation of one of the three Familiar hierarchical patterns. This was done to provide a baseline level of responsiveness following habituation as well as to check for regression to the mean. To ensure that this trial represented baseline responsiveness that was not unduly affected by one of the three hierarchical patterns presented during the habituation phase, infants in each of the two habituation groups at each age, respectively, were divided into three subgroups. Each of these subgroups was then tested with a different one of the three Familiar habituation-phase patterns. Following presentation of this initial Familiar test trial, each infant then received three additional test trials: the Familiar Non-Hierarchical test trial, the Generalization Hierarchical test trial and the Generalization Non-Hierarchical test trial. All of these three trials were presented in counterbalanced order across infants. Following the presentation of these three test trials, we presented the Winnie-the-Pooh cartoon again to check for possible fatigue effects and terminated the experiment.

2.2 | Results and discussion

In the first and preliminary analysis, we checked whether infants exhibited regression to the mean after they met the habituation criterion. This is a phenomenon that sometimes occurs in habituation studies where, for some inexplicable reason, infants sometimes exhibit response recovery to what is otherwise a familiar stimulus (Bertenthal, Haith, & Campos, 1983). To check for regression to the mean, we examined responsiveness during the Familiar Hierarchical test trial presented during the Familiar set of test trials. Any infant whose looking duration was more than two standard deviations above the group mean during this test trial was considered to have exhibited regression to the mean and that infant's data were not included in any subsequent analyses. Based on this analysis, we found that five 8–10 month-olds and five 12–14 month-olds met the criterion for regression to the mean and, as a result, were excluded from further analyses.

The mean duration-of-looking scores can be seen in Figure 3, plotted separately for the Familiar and Generalization test trials as a function of age. As can be seen, infants looked longer at the non-hierarchical than the hierarchical patterns in both types of test trials and at each age. We used a mixed, repeated-measures analysis of variance (ANOVA) to analyze the duration-of-looking scores, with Pattern Type (2; Hierarchical, Non-Hierarchical) and Test-Trial Type (2; Familiar, Generalization) as the within-subjects factors and Age (2; 8- to 10-month-olds, 12- to 14-month-olds) as a between-subjects factor. This analysis yielded several main effects, including a Pattern Type (F [1, 159] = 39.38, $p < .001$, $\eta_p^2 = .199$), Test-Trial Type (F [1, 159] = 235.52, $p < .001$, $\eta_p^2 = .597$), and an Age (1, 159) = 4.77, $p < .05$, $\eta_p^2 = .029$ effect. In addition, the analysis yielded a Test-Trial Type X Age interaction (F [1, 159] = 3.96, $p < .05$, $\eta_p^2 = .024$).

The Pattern Type effect indicates that, overall, infants looked longer in the non-hierarchical test trials (mean = 10.19 s, SD = 5.72) than in the hierarchical ones (mean = 6.78 s, SD = 4.33) and indicates that infants discriminated between the two types of patterns. The Test-Trial Type main effect indicates that infants looked longer in the Generalization test trials (mean = 11.9, SD = 6.1) than in the Familiar test trials (mean = 5.07, SD = 2.7) and demonstrates that infants detected the greater overall novelty of the Generalization patterns. The Age effect indicates that, overall, the 12- to 14-month-old infants looked longer (mean = 9.07 s, SD = 3.90) than did the 8–10 month-old infants (mean = 7.77 s, SD = 3.57). Finally, the Test-Trial Type X Age interaction indicates that responsiveness across the two types of test trials differed across age.

Despite the absence of a Pattern Type X Test-Trial Type X Age interaction, we investigated whether infants exhibited statistically significant response recovery to a non-hierarchical pattern relative to a hierarchical pattern in the Familiar and Generalization types of trials, separately. The decision to carry out these comparisons was based on our a priori theoretical expectations that infants would detect the difference between the two types of patterns and that it was possible that infants would detect them regardless of their age. Thus, we conducted separate analyses of the data from the Familiar and Generalization test trials, respectively, by way of mixed, repeated-measures ANOVAs, with Pattern Type (2) as the within-subjects factor and Age (2) as a between-subjects factor. For the Familiar set of test trials (Figure 3), we found a significant main effect of Pattern Type (F [1, 159] = 45.90, $p < .001$, $\eta_p^2 = .224$) and a significant Pattern Type X Age interaction (F [1, 159] = 5.35, $p < .05$, $\eta_p^2 = .033$). Planned comparisons indicated that each age group exhibited significant response recovery (t [71] = 3.05, $p < .01$, 2-tailed; t [88] = 6.71, $p < .001$, 2-tailed, respectively). For the Generalization set of test trials, we found a significant Pattern Type main effect (F [1, 159] = 15.90, $p < .001$, $\eta_p^2 = .091$). Planned comparisons indicated that each age group exhibited significant response recovery (t [71] = 3.27, $p < .01$; t [88] = 2.68, $p < .01$, respectively).

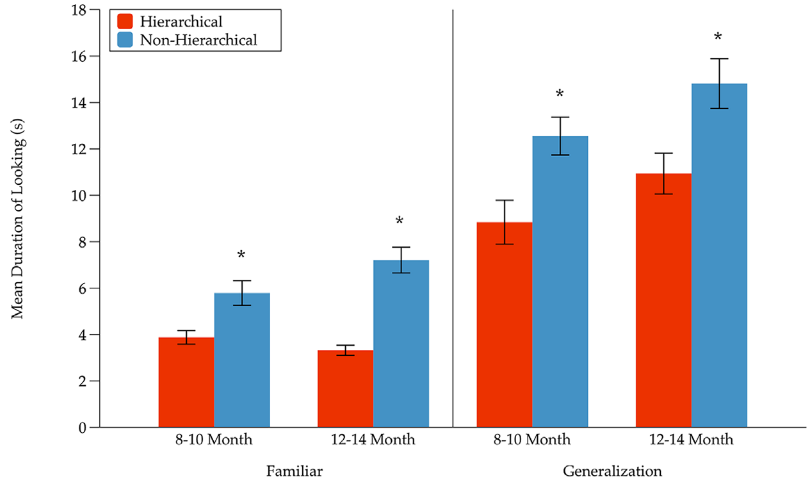

FIGURE 3 Mean duration of looking at the hierarchical and non-hierarchical patterns in the Familiar and Generalization test trials, respectively, in the 8- to 10- and 12- to 14-month-old infants in Experiment 1. Error bars indicate SEM and asterisks indicate a statistically significant difference in looking at the two types of patterns

Finally, we conducted an analysis to determine whether responsiveness to the different surface characteristics of hierarchical and non-hierarchical patterns might account for our results. In other words, we asked whether successful discrimination might have been due to something other than the detection of the underlying hierarchical structure per se. Although this issue is a notoriously difficult one to address (see Fountain, 2008, for a discussion of this problem in work with non-humans), one way of approaching it is to determine whether or not the hierarchical and non-hierarchical patterns differed systematically in terms of their surface characteristics in a way that might have provided a basis for their discrimination.

In general, hierarchical pattern sequences are locally more "regular" than non-hierarchical pattern sequences. For example, the hierarchical pattern sequence "22445533" appears to be more regular than its yoked non-hierarchical version "24324553." Of course, this perceived regularity could simply be another way of saying that the former is more (hierarchically) structured than the latter. Nonetheless, one way of characterizing this regularity is by examining the element-by-element interval structure of the two different pattern sequences. This can be accomplished with an interval content analysis of the pattern sequences, a procedure that has a long history in domains such as melodic contour processing (e.g., Friedmann, 1985, 1987; Marvin & Laprade, 1987; Quinn, 1999; Schmuckler, 1999, 2010). An interval content analysis simply computes the numerical difference between pattern sequence elements. In the present case, the difference corresponds directly to the spatial distance between the elements on the display screen. Such an analysis reveals that the hierarchical pattern sequence is likely to be more regular in that it potentially contains a smaller set of interval distances between elements. For example, for the hierarchical and non-hierarchical pattern sequences above, the interval content is <0 2 0 1 0 −2 0> and <2 −1 −1 2 1 0 −2>, respectively. Viewed in this light, the interval content of the hierarchical pattern sequence appears to generate smaller intervals, contain more repetitions of elements, and be less variable, than the non-hierarchical pattern sequence.

To quantify the level of surface information, we calculated multiple parameters of the interval content of these pattern sequences and compared these parameters across the hierarchical and non-hierarchical pattern sequences by way of *t*-tests. Specifically, the quantifications that we employed involved the summed absolute values of the intervals (representing the total amount of movement across the display, irrespective of left-right direction), the standard deviations of the signed interval values (representing the degree of back and forth movement across the display), a count of interval values of 0 (representing sequence repetition), and a count of interval values of ±1 (representing horizontal stepwise motion across the display). The results of these tests revealed that there were no significant differences for any of these parameters across the hierarchical and non-hierarchical pattern sequences. Therefore, given that traditional measures of surface features did not reveal any systematic differences across the two sets of sequences, the possible contribution of such features to the successful performance of the two oldest groups of infants can be ruled out.

In sum, the findings from this experiment are consistent with our first prediction, namely that starting at 8 months of age infants begin to exhibit evidence of hierarchical pattern learning. That is, we found that 8- to 10-month-old infants not only successfully learned specific hierarchical serial patterns but that they also successfully generalized their learning to novel hierarchical serial patterns.

3 | EXPERIMENT 2

Given the finding in Experiment 1 that infants exhibit successful learning and generalization of hierarchical serial patterns by 8- to 10-months of age, we then asked whether younger infants might be able to learn and generalize such patterns. As noted earlier, prior studies have found that relatively young infants generally only exhibit the simplest forms of pattern learning. For example, these studies have found that infants younger than 8 months of age can detect adjacent sequence statistics but that they do not detect ordinal position information nor simple rules (Frank, Slemmer, Marcus, & Johnson, 2009; Lewkowicz, 2013; Lewkowicz & Berent, 2009; Marcus et al., 1999). Therefore, we predicted that 4- to 6-month-old infants would not exhibit successful learning and generalization of the hierarchical serial patterns presented in Experiment 1.

3.1 | Method

3.1.1 | Participants

We tested forty-three 4- to 6-month-old infants ($M = 5.54$ months; SD = 1.17; 19 girls). We also tested an additional 13 infants but did not include their data due to health concerns ($n = 4$), fussiness ($n = 7$), inattentiveness ($n = 1$), and difficulty in coding looking behavior (n = 1).

3.1.2 | Apparatus and stimuli

The apparatus and stimuli were identical to those used in Experiment 1.

3.1.3 | Procedure

The procedure was identical to that used in Experiment 1.

3.2 | Results and discussion

The preliminary analysis of the data indicated that three 4–6 month-olds exhibited regression to the mean. These infants were excluded from any further analyses. The remaining data were analyzed with a repeated measures ANOVA, with Pattern Type (2; Hierarchical, Non-Hierarchical) and Test-Trial Type (2; Familiar, Generalization) as the within-subjects factors. The mean duration of looking scores for the Familiar and the Generalization test trials, respectively, appear in Figure 4. As can be seen, and consistent with our prediction, these infants did not exhibit differential looking in either set of test trials. This pattern was confirmed by the results of the ANOVA which indicated that neither the Pattern Type effect (F [1, 42] = 2.03, *ns*) nor the Pattern Type X Test-Trial Type interaction (F [1, 42] = 0.56, *ns*) were significant. The only significant effect was the Test Trial-Type effect (F [1, 42] = 15.36, $p < .001$, $\eta_p^2 = .268$) which, like in the older infants in Experiment 1, indicates that the novel attributes of both the hierarchical and non-hierarchical patterns presented in the Generalization set elicited greater overall attention.

3.2.1 | Direct comparisons of all three age groups

The data from Experiments 1 and 2 indicate that the ability to learn, discriminate, and generalize hierarchical serial patterns emerges between 4 and 6 and 8–10 months of age. To further confirm that this is the case, we conducted an analysis in which we compared the 3 age groups directly. To do so, we used a mixed, repeated-measures, ANOVA, with Pattern Type (2; Hierarchical, Non-Hierarchical) and Test-Trial Type (2; Familiar, Generalization) as the within-subjects factors and Age (3; 4–6, 8–10, and 12–14 month-olds) as the between-subjects factor. This analysis yielded a significant Test-Trial Type effect (F [1, 201] = 189.51, $p < .001$, $\eta_p^2 = .485$), a significant Pattern Type effect (F [1, 201] = 30.30, $p < .001$, $\eta_p^2 = .131$), a significant Test-Trial Type X Age interaction (F [2, 201] = 8.03, $p < .001$, $\eta_p^2 = .074$), and a significant Pattern Type X Age interaction (F [2, 201]= 3.19, $p < .05$, $\eta_p^2 = .031$).

The Test-Trial Type effect indicates that, together and regardless of age, infants responded more to the stimuli presented during the Generalization set of test trials than to the stimuli presented during the Familiar set of test trials. This main effect did, however, differ across age, mainly due to the fact that the two older groups of infants exhibited a greater response difference between the Familiar and Generalization test set than did the youngest group. These two effects indicate that greater responsiveness in the Generalization trials was due to the greater novelty of the stimuli in that trial set.

Of course, the most important result from the direct age comparisons is the statistically significant Pattern Type X Age

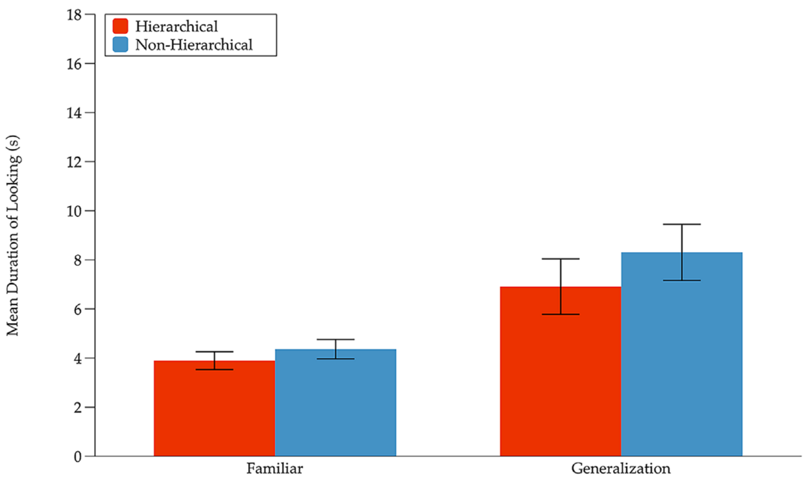

FIGURE 4 Mean duration of looking at the hierarchical and non-hierarchical patterns in the Familiar and Generalization test trials, respectively, in the 4- to 6-month-old infants in Experiment 2. Error bars indicate SEM

interaction. This interaction indicates that responsiveness in the test trials differed across the 3 ages. The locus of that difference is between the youngest group of infants and the two oldest groups. Whereas the 4- to 6-month-old infants exhibited no evidence of successful learning and discrimination in neither the Familiarization nor Generalization trials, the 8- to 10-month-old and the 12- to 14-month-old did exhibit such evidence. Together, these findings show clearly that the ability to learn, discriminate, and generalize hierarchical serial patterns emerges during infancy.

4 | GENERAL DISCUSSION

Pattern perception and learning are fundamental processes that are implicated across most major domains of psychological functioning including speech, language, music, event perception, motor action, and social interaction. Given this, it would be highly adaptive if these abilities emerged early in development. To examine whether this might be the case, we investigated the perception, learning, discrimination, and generalization of rule-bound hierarchical serial patterns in infancy by testing 4- to 6-, 8- to 10-, and 12- to 14-month-old infants. We found no evidence of learning and generalization of hierarchical serial patterns in 4–6 month-old infants but did find evidence of it in 8–10 and 12–14 month-old infants.

Our findings are fully consistent with the findings from the only other study that has investigated learning of hierarchical rules in infancy (Werchan et al., 2015). This study showed that 8-month-old infants can learn hierarchical rules that specify associations between distinct object properties such as shape, color, and location or that specify unique associations between different people vocalizing specific sounds, objects, and words and that they can then use these rules for making corresponding oculomotor responses. The findings from the current study add to the results from the Werchan et al. (2015) study by demonstrating that infants also can perceive, learn, discriminate, and generalize hierarchical serial patterns. In the aggregate, the findings show that sensitivity to hierarchical structure across a range of instantiations, from stimulus categorization to spatiotemporally organized serial patterns, emerges during infancy.

As noted in the Introduction, infants develop in a spatiotemporally organized world that is full of discrete events whose specific spatiotemporal structure defines their meanings. Therefore, one of the infant's tasks is to perceive, extract, and learn such structure in order to behave appropriately. Here, we asked whether and when infants might become capable of performing such a task by using a variation of the habituation/test procedure. This multiple habituation procedure makes it possible to ask two distinct questions. One is whether infants can learn and discriminate the specific attributes of different instantiations of a particular type of event. The other is whether infants can also extract the invariant property that defines the different instantiations of the event and then apply this knowledge to the discrimination of novel instances of a similar event. Hence, first we habituated infants to three different hierarchical serial patterns, each defined by a different hierarchical rule. This provided the infants with the opportunity to learn the specific rules for each pattern and also the general concept of hierarchical organization. Then, we administered two types of test trials (Familiar and Generalization) to determine whether the infants learned the specific patterns, whether they extracted the general concept of hierarchical organization, and whether they could generalize this concept to the discrimination of novel hierarchical serial patterns. During the Familiar test trials, we contrasted responsiveness to one of the three hierarchical serial patterns presented during the habituation phase versus responsiveness to its yoked non-hierarchical version. During the Generalization test trials, we contrasted responsiveness to a novel hierarchical serial pattern versus responsiveness to the same but non-hierarchical version of this pattern. The finding that the two oldest age groups exhibited successful discrimination in both the Familiar and Generalization test trials, but that the 4- to 6-month-old infants did not exhibit successful discrimination in either type of test trials, demonstrates that the ability to perceive and learn the concept of hierarchical serial organization emerges by 8–10 months of age.

Of course, our interpretation of these findings rests on the assumption that simpler, alternative explanations can be ruled out. Two alternative explanations—one based on methodological and the other on conceptual considerations—can be discounted. In terms of methodological considerations, the yoking procedure that we used to generate the non-hierarchical versions of the hierarchical patterns ensured that the identical sequence elements were activated across the two types of patterns. Consequently, the temporal aspects of the two types of patterns were not confounded by spatial difference cues. Put differently, the only difference between the two types of patterns was that they were activated in temporally different ways. This, in turn, indicates that the 8- to 10- and 12- to 14-month-old infants' successful discrimination could only have been based on the detection of serial pattern differences, and not on spatial differences (e.g., variation in the number of active elements in any given part of the display as a function of hierarchical versus non-hierarchical patterns). Also, in terms of methodological considerations, the use of identical pattern elements and the presentation of multiple exemplars of hierarchical serial patterns during the habituation phase ensured that pattern statistics did not mediate successful discrimination in the Familiar and Generalization test trials at the two older ages. This is because: (a) computation of the statistics of patterns composed of identical elements is very difficult; (b) there is no evidence that infants can compute the statistics of patterns composed of identical elements; (c) any adjacent-element statistics could not have mediated responsiveness because the specific ordinal-position statistics varied across the habituation phase; and (d) the statistics in the Generalization test trials were novel.

Although the surface analysis indicated that the patterns presented in the current study did not differ in terms of their surface characteristics, there is one other aspect of hierarchical patterns that has concerned researchers in this area. In essence, it has frequently been noted that hierarchical patterns are characterized by a fundamental ambiguity in that they can potentially be represented in multiple ways (Fountain, 2008; Fountain & Doyle, 2011; Greeno & Simon, 1974; Jones, 1981). This issue has been most elegantly addressed by Fountain and colleagues (Fountain, 2008; Fountain & Doyle, 2011; Fountain & Rowan, 1995; Garlick et al., 2017; Muller & Fountain, 2016) who have provided compelling evidence for the existence of hierarchical versus associative representations in animals. For our purposes, the critical aspect is not the exact nature of the hierarchical representation, but rather that our findings demonstrate that infants can reliably discriminate hierarchically structured from non-hierarchically (e.g., linear or associative) patterns. Moreover, the discrimination of novel hierarchical from non-hierarchical patterns also demonstrates that infants can abstract out the concept of hierarchical structure per se and that they can generalize this concept to the perception of patterns that they have not previously experienced. This shows that their successful discrimination of novel hierarchical from non-hierarchical serial patterns could not have been based on the associative structure of the patterns.

It is interesting to note that our demonstration of the emergence of a sensitivity to hierarchical structure by 8 months of age converges strongly with the findings from the Werchan et al. (2015) study. Whereas Werchan et al. (2015) demonstrated that 8-month-old infants are sensitive to the hierarchical structuring of rules used for categorization of objects and for guiding subsequent responsiveness, we showed that starting at this same age infants can also perceive and learn the hierarchical structure of spatiotemporally extended events. Earlier, we suggested the possibility that sensitivity to the hierarchical structure of serial patterns may emerge later, given its explicit reliance on the ability to abstract and integrate information across space and time over complex stimulus arrays. This possibility can be ruled out given the fact that the findings from the Werchan et al. (2015) study and the current one indicate that sensitivity to hierarchical information in static and dynamic contexts is present by 8 months of age. Such convergence leads to the conclusion that the sensitivity to hierarchical structure that is present by 8 months of age reflects the operation of domain-general perceptual and learning mechanisms.

On the one hand, the emergence of such a sophisticated ability this early in development may seem somewhat surprising. On the other hand, this is less surprising when it is recognized that by the time this sensitivity emerges, infants have had 8 months of exposure to hierarchical organization in speech, music, object categorization, social interaction, and their own motor action. Indeed, the combined findings from the Werchan et al. (2015) study and from the current one are also important from a general theoretical perspective because they demonstrate that the ability to learn hierarchical structure is present prior to the emergence of language and the emergence of other key cognitive skills where detection of hierarchical structure is critical (Greenfield, 1991; Hauser et al., 2002; Jackendoff & Pinker, 2005; Zacks & Tversky, 2001). This makes it likely that this early emerging skill is one critical component of the perceptual foundation that is essential to the development of higher-level cognitive and motor functions.

The conclusion that the aggregate findings to date reflect the operation of a domain-general mechanism are further bolstered by the fact that the hierarchical serial patterns presented in the current study were comprised of abstract and identical objects together with synchronous non-speech sounds. Therefore, the infants in the current study could not have relied on any domain-specific knowledge to successfully perceive and learn the hierarchical rules. The domain-general interpretation offered here is also consistent with evidence that pattern learning in infancy is mediated by general-purpose mechanisms (Saffran et al., 2007) and that the mechanisms required for the learning of linguistic and/or musical syntax do not emerge from domain-specific mechanisms but, rather, from general pattern detection mechanisms (Saffran & Thiessen, 2007; Shafto, Conway, Field, & Houston, 2012). Therefore, given that the general-purpose hierarchical pattern learning mechanisms found here are functional prior to the emergence of frank language-processing mechanisms, it is reasonable to conclude that the sort of skill uncovered here is likely to contribute to cognitive development in specific domains.

The findings on infant perception and learning of hierarchical serial patterns as well as findings on infants' ability to perceive and learn pattern statistics, rhythmic/prosodic structure, the ordinal position of sequence elements, and simple reduplication rules raise some additional theoretical questions. The main one is about the role of early experience. We have already suggested that early experience is likely to contribute to the emergence of pattern perception and learning skills because infants have ample exposure to patterned events. These consist of social partners' speech and actions, the spatiotemporally distributed actions of non-social objects, and their own, self-generated actions. Social partners are known to scaffold the behaviors that they direct to infants based on the ability to respond to them and, thus, increase the complexity of their behaviors as infants grow. In terms of self-generated patterned behaviors, infants also produce increasingly more complex patterned behaviors, that include rhythmical sucking, rhythmical hand waving, and banging, babbling, and sequencing of phonemes to form words, as they grow (Iverson & Thelen, 1999). This is interesting because it is known that these various rhythmical actions are generated by biological oscillators that dynamically couple together and that, ultimately, lead to a tight neural and functional coordination between emerging motor gestures, speech production, and language (Iverson & Goldin-Meadow, 2005; Iverson & Thelen, 1999; Tzeng & Wang, 1984). It is also known that, at the neural level, this gesture-speech-language coupling is mediated by a precise timing/sequencing mechanism that permits the decoding and generation of complex serial patterns (Ojemann, 1984). Given that gestures, speech, language, and increasingly more sophisticated perceptual processing skills all emerge gradually during infancy, and given that action and perception are tightly linked in infancy (Gibson, 1988; Schmuckler, 1993; Schmuckler, 2013; Thelen & Smith, 1994), it is theoretically reasonable to posit that the underlying structural and functional architecture is likely to begin supporting the perception and learning of hierarchical serial patterns in infancy too. Certainly, the findings from the Werchan et al. (2015) study and our study are consistent with this theoretical scenario. Moreover, findings from a follow-up study by Werchan, Collins, Frank and Amso (2016) have shown prefrontal cortex involvement and striatal involvement using eye blink rate in infant learning of the hierarchical rules.

Finally, our findings that the older infants learned the concept of recursive hierarchical organization from patterns specified by spatiotemporally extended audiovisual information are noteworthy because this is the kind of information that specifies most social communication events. For example, social partners usually either communicate with spatiotemporally distributed gestures and corresponding vocalizations or with spatiotemporally distributed audible and visible speech utterances. In either case, if one of the partners is to extract meaningful information from such gestures or utterances, he or she must be able to perceive their hierarchical spatiotemporal structure. Interestingly, infants begin to lipread (Hillairet de Boisferon, Tift, Minar, & Lewkowicz, 2017; Lewkowicz & Hansen-Tift, 2012; Pons, Bosch, & Lewkowicz, 2015) at the same point in development that they begin to exhibit the ability to learn complex hierarchical sequential patterns. This means that by the time infants are 8–10 months of age, they can take advantage of, both, the highly salient audiovisual speech and language cues available in their interlocutor's mouth as well as the spatiotemporally patterned cues that specify their interlocutors' utterances. There is little doubt that the co-emergence of these two skills is likely to facilitate the acquisition of speech, language, and other general communicative and social interaction skills.

ACKNOWLEDGMENTS

Supported by the National Science Foundation (Grant BCS-0751888) and by the Eunice Kennedy Shriver National Institute of Child Health & Human Development (Grant R01HD057116) to D. J. L. and by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada to M. A. S. We thank Kelly Henning for her assistance.

ORCID

David J. Lewkowicz http://orcid.org/0000-0001-6685-3178

REFERENCES

Amso, D., Haas, S., McShane, L., & Badre, D. (2014). Working memory updating and the development of rule-guided behavior. *Cognition*, *133*, 201–210.

Badre, D. (2008). Cognitive control, heirarchy, and the rostro-caudal axis of the prefrontal cortext. *Trends in Cognitive Science*, *12*, 193–200.

Badre, D., & D'Esposito, M. (2007). FMRI evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience*, *19*, 1–18.

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron*, *66*, 315–326.

Bertenthal, B. I., Haith, M. M., & Campos, J. J. (1983). The partial-lag design: A method for controlling spontaneous regression in the infant-control habituation paradigm. *Infant Behavior and Development*, *6*(3), 331–338.

Chang, H. W., & Trehub, S. E. (1977). Infants' perception of temporal grouping in auditory patterns. *Child Development*, *48*(4), 1666–1670.

Cohen, L. B. (2009). The evolution of infant cognition: A personal account. *Infancy*, *14*(4), 403–413.

Cohen, L. B., & Strauss, M. S. (1979). Concept acquisition in the human infant. *Child Development*, *50*(2), 419–424.

Demany, L., McKenzie, B., & Vurpillot, E. (1977). Rhythm perception in early infancy. *Nature*, *266*(5604), 718–719.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826.

Fountain, S. B. (2008). Pattern structure and rule induction in sequential learning. *Comparative Cognition and Behavior Reviews*, *3*, 66–85. https://doi.org/10.3819/ccbr.2008.30004.

Fountain, S. B., & Doyle, K. E. (2011). Association and abstration in sequential learning: "What is learned?" revisited. *International Journal of Comparative Psychology*, *24*, 437–459.

Fountain, S. B., & Rowan, J. D. (1995). Coding of hierarchical versus linear pattern structure in rats and humans. *Journal of Experimental Psychology: Animal Behavior Processes*, *21*, 187–202.

Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. *Developmental Science*, *12*, 504–509.

Friedmann, M. L. (1985). A methodology for the discussion of contour: Its application to Schoenberg's music. *Journal of Music Theory*, *29*, 223–248.

Friedmann, M. L. (1987). My contour, their contour. *Journal of Music Theory*, *31*, 268–274.

Garlick, D., Fountain, S. B., & Blaisdell, A. P. (2017). Serial pattern learning in pigeons: Rule-based or assoiative? *Journal of Experimental Psychology: Animal Learning and Cognition*, *43*, 30–47.

Gerken, L. (2006). Decisions, decisions: Infant language learning when multiple generalizations are possible. *Cognition*, *98*(3), B67.

Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology*, *39*, 1–41.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Gómez, R. L., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206.

Greenfield, P. M. (1991). Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and Brain Sciences*, *14*(4), 531–595.

Greeno, J. G., & Simon, H. A. (1974). Processes for sequence production. *Psychological Review*, *81*, 187–198.

Hauser, M. D., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, *298*(5598), 1569–1579.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2017). Selective attention to a talker's mouth in infancy: role of audiovisual temporal synchrony and linguistic experience. *Developmental Science*, *20*(3), https://doi.org/10.1111/desc.12381.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological Science*, *16*(5), 367–371.

Iverson, J. M., & Thelen, E. (1999). Hand, mouth and brain. The dynamic emergence of speech and gesture. *Journal of Consciousness Studies*, *6*(11-12), 19–40.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition*, *97*(2), 211–225.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–345.

Jones, M. R. (1981). A tutorial on some issues and methods in serial pattern research. *Perception and Psychophysics*, *30*, 492–504.

Kayser, A. S., & D'Esposito, M. (2013). Abstract rule learning: The differential effects of lesions in frontal cortex. *Cerebral Cortex*, *23*, 230–240.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42.

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, location, location: Development of spatiotemporal sequence learning in infancy. *Child Development*, *78*(5), 1559–1571.

Kovotsky, K., & Simon, H. A. (1973). Empirical tests of a thoeyr of human acquisition of concepts for sequential patterns. *Cognitive Psychology*, *4*, 399–424.

Lashley, K. S., (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium*. (pp. 123–147). New York: Wiley.

Lewkowicz, D. J. (2013). Development of ordinal sequence perception in infancy. *Developmental Science*, *16*(3), 352–364.

Lewkowicz, D. J., & Berent, I. (2009). Sequence learning in 4 month-old infants: Do infants represent ordinal information? *Child Development*, *80*(6), 1811–1823.

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436.

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4- to 10-month-old infants. *Developmental Psychobiology*, *48*, 288–300.

Mandel, D. R., Nelson, D. G., & Jusczyk, P. W. (1996). Infants remember the order of words in a spoken sentence. *Cognitive Development*, *11*(2), 181–196.

Marcovitch, S., & Lewkowicz, D. J. (2009). Sequence learning in infancy: The independent contributions of conditional probability and pair frequency information. *Developmental Science*, *12*(6), 1020–1025.

Marcus, G. F., Vijayan, S., Rao, S., & Vishton, P. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, *79*, 487–509.

Marvin, E. W., & Laprade, P. A. (1987). Relating musical contours: Extensions of a theory for contour. *Journal of Music Theory*, *31*, 225–267.

Muller, M. D., & Fountain, S. B. (2016). Concurrent cognitive processes in rat serial learning: II. Discrimination learning, rule learning, chunk length, and multiple-item memories. *Journal of the Experimental Analysis of Behavior*, *105*, 155–175.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception & Performance*, *24*(3), 756–766.

Ojemann, G. A. (1984). Common cortical and thalamic mechanisms for language and motor functions. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *246*(6), R901–R903.

Pickens, J., & Bahrick, L. E. (1997). Do infants perceive invariant tempo and rhythm in auditory-visual events? *Infant Behavior and Development*, *20*, 349–357.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science*, *26*(4), 490–498.

Quinn, I. (1999). The combinatorial model of pitch contour. *Music Perception*, *16*, 439–456.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review*, *77*(6), 481–495.

Restle, F. (1972). Serial patterns: The role of phrasing. *Journal of Experimental Psychology*, *92*, (385–390).

Restle, F., & Brown, E., (1970). Organization of serial pattern learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory (Vol. 4)*. (pp. 249–332). New York: Academic Press.

Restle, F., & Burnside, B. L. (1972). Tracking of serial patterns. *Journal of Experimental Psychology*, *95*(2), 299.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

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*(1), 27–52.

Saffran, J. R., Pollak, S. D., Seibel, R. L., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*(3), 669–680.

Saffran, J. R., & Thiessen, E. D., (2007). Domain general learning capacities. In E. Hoff, & M. Schatz (Eds.), *Blackwell handbook of language development*. (pp. 68–86). Oxford, UK: Blackwell Publishing.

Schmuckler, M. A., (1993). Perception-action coupling in infancy. In G. J. P. Savelsbergh (Ed.), *The development of coordination in infancy*. (pp. 137–173). North-Holland: Elsevier.

Schmuckler, M. A. (1999). Testing models of melodic contour similarity. *Music Perception*, *16*, 295–326.

Schmuckler, M. A. (2010). Melodic contour similarity using folk melodies. *Music Perception*, *28*(2), 169–193. https://doi.org/10.1525/mp.2010.28.2.169.

Schmuckler, M. A., (2013). Perceptual-motor relations in obvious and non-obvious domains: A history and review. In P. D. Zelazo (Ed.), *Handbook of developmental psychology*. (pp. 237–270). Oxford, UK: Oxford University Press.

Shafto, C. L., Conway, C. M., Field, S. L., & Houston, D. M. (2012). Visual sequence learning in infancy: Domain-General and Domain-Specific associations with language. *Infancy*, *17*(3), 247–271.

Simon, H. A. (1972). Complexity and the representation of patterned sequences of symbols. *Psychological Review*, *79*, 369–382.

Simon, H. A., & Kotovsky, H. (1963). Human acquisition of concepts for sequential patterns. *Psychological Review*, *70*, 534–546.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

Tzeng, O., & Wang, W. (1984). Search for a common neurocognitive mechanism for language and movements. *American Journal of Physiology-Regulatory, Integrative, and Comparative Physiology*, *246*(6), R904–R911.

Unger, K., Ackerman, L., Chatham, C. H., Amso, D., & Badre, D. (2016). Working memory gating mechanisms explain developmental change in rule-guided behavior. *Cognition*, *155*, 8–22.

Vitz, P. C., & Todd, T. C. (1969). A coded element model of the perceptual processing of sequential stimuli. *Psychological Review*, *76*, 433–449.

Werchan, D. M., Collins, A. G., Frank, M. J., & Amso, D. (2015). 8-month-old infants spontaneously learn and generalize hierarchical rules. *Psychological Science*, *26*(6), 805–815. https://doi.org/10.1177/0956797615571442

Werchan, D. M., Collins, A. G., Frank, M. J., & Amso, D. (2016). Role of prefrontal cortex in learning and generalizing hierarchical rules in 8-month-old infants. *Journal of Neuroscience*, *36*(40), 10314–10322.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Lewkowicz DJ, Schmuckler MA, Mangalindan DMJ. Learning of hierarchical serial patterns emerges in infancy. *Developmental Psychobiology*. 2018;1–13. https://doi.org/10.1002/dev.21614