Visual statistical learning in infancy: Discrimination of fine-grained regularities depends on early test trials

Julie Bertels1,2 | Estibaliz San Anton1 | Emeline Boursain1 | Hermann Bulf3,4 | Arnaud Destrebecqz1

1ULBabyLab, Consciousness, Cognition, and Computation Group (CO3), Center for Research in Cognition and Neurosciences (CRCN), Université libre de Bruxelles, Brussels, Belgium
2Laboratoire de Cartographie Fonctionnelle du Cerveau (LCFC), ULB Neuroscience Institute (UNI), Université libre de Bruxelles, Brussels, Belgium
3Dipartimento di Psicologia, Università degli Studi di Milano – Bicocca, Milano, Italy
4NeuroMI – Milan Center for Neuroscience, Milano, Italy

Abstract
Infants’ ability to detect statistical regularities between visual objects has been demonstrated in previous studies (e.g., Kirkham et al., Cognition, 83, 2002, B35). The extent to which infants extract and learn the actual values of the transitional probabilities (TPs) between these objects nevertheless remains an open question. In three experiments providing identical learning conditions but contrasting different types of sequences at test, we examined 8-month-old infants’ ability to discriminate between familiar sequences involving high or low values of TPs, and new sequences that involved null TPs. Results showed that infants discriminate between these three types of sequences, supporting the existence of a statistical learning mechanism by which infants extract fine-grained statistical information from a stream of visual stimuli. Interestingly, the expression of this statistical knowledge varied between experiments and specifically depended on the nature of the first two test trials. We argue that the predictability of this early test arrangement—namely whether the first two test items were either predictable or unexpected based on the habituation phase—determined infants’ looking behaviors.
A fundamental aspect of human cognition is our ability to automatically and incidentally detect recurring patterns, namely statistical regularities, in our sensory environment. Usually referred to as “statistical learning” (hereafter, SL), this ability has become a major topic in cognitive science in the past 25 years. In particular, many studies have documented the infants’ ability to extract regularities in their environment and have shown that SL is a central learning mechanism in infancy.

Research on this major topic started in the auditory modality. In a seminal study, Saffran et al. (1996) showed that 8-month-old infants are able to extract the statistical information within a speech stream based solely on the statistical relationships between speech sounds (see also, Aslin et al., 1998), a skill that contributes to language acquisition (Romberg & Saffran, 2010). These studies were promptly followed by others examining infants’ ability to detect and extract regular patterns in the visual modality. Regularities are indeed ubiquitous across space and time in the visual environment: They exist in both how objects are arrayed in space and how they appear over time (Turk-Browne, 2012). Yet, capturing these regularities would help the infant organizing and integrating his/her visual experience into coherent representations of the world. Accordingly, further studies have demonstrated that infants can detect regularities between visual objects, presented either together (Bertels et al., 2016; Fiser & Aslin, 2002; Wu et al., 2011) or sequentially (Kirkham et al., 2002), as well as between continuous, dynamic visual events such as in sequences of actions (Roseberry et al., 2011; Stahl et al., 2014).

Kirkham et al. (2002) were the first to show SL in the visual domain. In that study, infants were exposed to a continuous stream of three randomly presented pairs of colored shapes, a pair consisting of two shapes presented successively. Since there were no pauses between pairs, the only way to extract regularities from the visual stream was to rely on the transitional probabilities (hereafter, TPs)\(^1\) between shapes that were higher within (TP = 1) than between pairs (TP = .33). Once the infant was habituated to the stream, sensitivity to the between-shapes TPs was assessed by measuring infants’ looking behavior while they were presented with “familiar” sequences, made up of the same three pairs of shapes as in the habituation phase, and with “novel” sequences, made up of the random presentation of the same six shapes. Results showed that infants at 2, 5, and 8 months of age looked longer at novel than familiar sequences, indicating their ability to discriminate between structured versus unstructured visual patterns.

Many studies have replicated these findings since then, using different types of visual stimuli (Emberson et al., 2019), in younger (Bulf et al., 2011) and older infants (Antovich et al., 2020). Critically, these findings have also been replicated using test designs that allowed concluding more straightforwardly that infants were sensitive to the statistical relations between elements (i.e., TPs and/or co-occurrence frequency), rather than merely to local redundancies or frequency of individual elements (Marcovitch & Lewkowicz, 2009; Slone & Johnson, 2015). In these studies, instead of continuous test sequences including all elements (either presented by pairs in the familiar sequences, or randomly in the novel ones), each test trial consisted of the repeated presentation of a single sequence unit (a pair or a triplet, depending on the studies), in a familiar or unfamiliar arrangement of the constituting elements. Doing so, test trials only differed by the TPs between sequence elements: high (i.e., TP = 1) in familiar trials or low (i.e., TP < 1) in unfamiliar ones (Marcovitch & Lewkowicz, 2009; Roseberry et al., 2011; Slone & Johnson, 2015; Stahl et al., 2014; Tummeltshammer et al., 2016). Although this design probably made the task even more challenging because it requires segmenting

\(^1\)The transitional probability is the probability of occurrence of a shape B after a shape A, i.e. \(P(B/A)\).
the stream based on the TPs rather than simply differentiating structured from random sequences, infants discriminated between familiar and unfamiliar trials. These results thus support that infants extracted and learned the statistical relations between sequence elements.

While infants’ ability to learn sequential regularities between visual stimuli is thus well established, their level of sensitivity to TP values and, therefore, the level of precision of their ability to discriminate between familiar and less familiar element units remain largely unexplored. Indeed, the aforementioned studies only contrasted infants’ looking durations to familiar, high TPs test pairs (i.e., with TP = 1) and one single level of lower TPs test pairs (e.g., TP = .5 or TP = .33, but see Tummeltshammer & Kirkham, 2013). Hence, one may ask whether this discrimination reflects that infants merely learn pairs or triplets of sequence elements that are bound together by high TPs and consider low TPs test sequences as unfamiliar, or whether they consider low TPs test sequences as familiar, but differing from high TPs test sequences in terms of the strength of the statistical relation between the constituting elements.

Slone and Johnson (2018) recently addressed a very similar topic. These authors questioned whether infants truly encode statistical relations between elements, as posited by statistical models, or whether they rather encode statistically coherent units of information from the input (i.e., “clusters” or “chunks”), hence disregarding the actual value of the statistical association between individual elements, as posited by chunking models. According to statistical models, infants perform precise statistical computations when exposed to a stream of stimuli (Aslin & Newport, 2012) and store the overarching statistical structure of the sequence in memory. Alternatively, according to chunking models, infants’ sensitivity to statistical regularities results from the extraction of individual elements that frequently follow each other, and from the storage of these elements in memory as individual “chunks” in an all-or-nothing fashion (i.e., sequential items in the input are either chunked or not), which are recalled as units when necessary (Perruchet & Peereman, 2004; see PARSER, Perruchet & Vinter, 1998, or TRACX, French et al., 2011). Slone and Johnson (2018) convincingly demonstrated that infants’ SL performance is best accounted for by chunking models, echoing recent connectionist work dedicated to modeling infant visual statistical learning (TRACX2; Mareschal & French, 2017). Nevertheless, as Slone and Johnson acknowledged, their data do not rule out the possibility that infants compute statistics in the process of chunk formation. The extraction of statistical regularities from the input could indeed be a step toward the building of chunks. Furthermore, previous studies have shown that different SL mechanisms might coexist and that their recruitment would depend on the nature of the input (Tummeltshammer et al., 2016) or on the learning condition (Franco & Destrebecqz, 2012). Hence, even if chunk-based representations arise from exposure to statistical regularities under certain circumstances, this would not preclude infants to encode statistical relations between elements (see also Endress et al., 2020). In the present study, we aimed at examining whether infants are able to make fine-grained distinctions between sequences involving various levels of TPs. This would indicate that their representations of the relations between visual sequential elements include the value of the transitional probabilities between them.

Encoding precise statistical information about the strength with which a visual element predicts the next one indeed has a high adaptive value, considering that the infant can use this knowledge to modify his/her behavior. Anticipating or planning what is coming next and how likely, with the anticipatory strength depending on the strength of the TPs, could indeed help the infant to act adequately and promptly. As a matter of fact, previous studies have shown that infants make use of their statistical knowledge to plan saccades in predictable locations (Kirkham et al., 2007) or to identify the action’s endpoints in a sequence of movements (Baldwin et al., 2008). Obviously, considering the strength of the TPs between visual events to anticipate their occurrence and modifying his/her behavior accordingly would be highly beneficial for the infant. Furthermore, such
graded behavior could in turn influence the infant’s statistical learning process. Infants would then fine-tune their knowledge of the strength of the TPs, depending on the adequacy of their behavior.

In three experiments providing identical learning conditions (i.e., we used infant-controlled habituations with the very same pairs of stimuli), we contrasted infants’ looking times to looped presentations of test pairs implementing high, low, and null TPs (i.e., differing by the strength of the TPs between visual elements). High TPs sequences were the most familiar as they were made up of the repeated presentation of one pair presented during habituation. Low TPs sequences were familiar too, although to a lesser extent: They were made up of the repeated presentation of the last element of a pair followed by the first element of another one. Finally, null TPs sequences were completely unfamiliar as they never occurred during habituation: they consisted of the repeated presentation of the first (or last) element of a pair and the first (or last) element of another pair.

In the first two experiments, we implemented the standard testing procedure in which looking times are compared between two different values of TPs. In Experiment 1, we aimed at replicating and extending past work demonstrating that infants are able to discriminate between familiar pairs differing by the strength of the transitional probabilities between their constituting elements (i.e., High vs. Low TPs pairs). In Experiment 2, we questioned whether Low TPs pairs are indeed considered as familiar and contrasted infants’ looking times to these and completely novel pairs with null TPs. Finally, in Experiment 3, we aimed at comparing infants’ interest to these three types of pairs, in a within-subject design in which the three TPs values were presented in a single test phase.

We expected that infants would differentially attend to high and low TPs sequences, as reported in previous studies (Slone & Johnson, 2015; Tummelshammer et al., 2016). Critically, if infants extract and store precise statistical information from the visual stream as statistical models predict, we expect them to discriminate between low and null TPs sequences. Rather, if they only learn to associate sequence elements bound by strong TPs into chunks, their looking times should not differ between low and null TPs test sequences. Finally, we expected infants to discriminate between high and null TPs sequences, since the difference in terms of TPs is even larger than between high and low TPs sequences.

2  |  EXPERIMENT 1: HIGH VERSUS LOW TPs

2.1  |  Method

2.1.1  |  Participants

Twenty-one full-term 7- to 9-month-old infants (13 females) composed the final sample (M age = 247 days, SD = 22, range = 214–300). Ten additional infants were tested but excluded from the analyses due to fussiness (n = 8), or failure to meet the habituation criterion (n = 2). This sample size provides a power of .87, given a .05 significance criterion in a one-tailed test and an effect size of .62 (as reported by Slone & Johnson, 2015, in the older group of infants).

This corresponds to a dropout rate of 32%, which is comparable to those reported in studies using similar stimuli and procedure. Except for Kirkham et al. (2002) reporting a dropout rate of only 16% (9 infants discarded out of 57 tested), more recent studies report much higher dropout rates. Marcovitch and Lewkowicz (2009) discarded 79 infants on a total of 175 tested (dropout rate = 45%). Addyman and Mareschal (2013) excluded the data of 21 infants on a total of 58 tested (dropout rate = 36%). Slone and Johnson (2015) tested 61 infants and only analyzed the data of 40 of them (dropout rate = 33%).
All infants had no known developmental difficulties. The infants were recruited via announcements in nurseries, in maternity hospitals, in the University newsletter, and on social networks. Infants received a gift (a book or a soft toy) and a diploma for their participation. The present study was conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent for each child before any assessment or data collection. All experimental protocols were approved by the Ethics Committee of the Psychological and Educational Sciences Faculty at the Université Libre de Bruxelles. The methods were carried out in accordance with the approved guidelines.

2.1.2 | Apparatus

Stimulus presentation, timing, and data collection were controlled using the Habit X 1.0 software (Cohen et al., 2004) running on a Mac mini. Stimuli were presented through a 55” LCD digital TV screen that was connected to the computer. The infants’ looking behavior was monitored through a video camera located above the TV screen and connected to a Dell computer.

2.1.3 | Stimuli

The stimuli were the same as those used in Kirkham et al. (2002), namely six colored shapes presented on a gray background, grouped in three pairs (i.e., sequences of two shapes, see Figure 1a). The same pairs were presented to all infants. Each shape was presented for one second and loomed from 10 to 41 cm in height (3.8°–15.3°).

2.1.4 | Procedure

Each infant was tested individually while seated on their parent’s lap 150 cm from the TV screen. The testing took place inside a 3-sided enclosure, in a dimly lit room. The experimenter initiated the trials and coded the infant’s looking behavior by using the computer keyboard. She was kept unaware of the identity of the stimuli presented to the infant.

The experiment was divided into two phases: a habituation and a test phase. During both phases, each trial began with the presentation of an attention getter (a looming star) presented in the middle of the screen concurrently with an attractive sound. As soon as the infant looked at the star, the experimenter initiated the trial.

Each habituation trial consisted of the repeated presentation of the three pairs in a pseudo-random order (maximum two successive presentations of a pair), one shape at a time in a continuous stream with no break between shapes (Figure 1a). Consequently, the transitional probability (TP) within pairs was 1, while the TP between pairs was .33. A habituation trial lasted a maximum of 60 s or until the infant’s gaze had shifted away from the screen for two consecutive seconds. When one of these criteria was met, the duration of the trial (i.e., the infant’s looking time) was recorded, and the next trial began by the presentation of the attention getter. The habituation phase ended once the cumulative looking time on four consecutive trials decreased to a value <50% of the cumulative looking time for the first four habituation trials (Kirkham et al., 2002; Marcovitch & Lewkowicz, 2009). If this criterion was not met after 12 trials, the habituation was considered unsuccessful and the infant’s data were excluded from the analyses (Slone & Johnson, 2015).
The test phase immediately followed the habituation phase, as soon as the habituation criterion had been reached. On each test trial, a pair of shapes was presented repeatedly, one 1-s looming shape at a time, with no break between each pair presentation, as in the habituation phase. A test trial lasted a maximum of 60 s or until the infant’s gaze was directed away from the screen for two consecutive seconds.

There were six test trials from two different conditions differing according to the TPs between shapes: High and Low TPs (3 trials per condition, chosen randomly for each infant among all possible pairs in one condition). In the High TPs sequences (Figure 1b), a pair from the habituation phase was presented in a loop, so that TPs from one shape to another were alternating between 1 and .33. In the Low TPs sequences (Figure 1c), the pair was made up of a shape appearing in the first position of a pair from the habituation phase and of a shape appearing in the second position of another pair from the habituation phase, so that TPs from one shape to another were alternating between 0 and .33. Test trials were presented pseudo-randomly so that a single type of test trial (High or Low TPs) was not presented more than twice in a row. Mean looking times across each of the two conditions were computed.

2.2 | Results

Infants were presented with 6–12 habituation trials (M = 8.048, SD = 2.179). Mean duration of the habituation phase was 206 s (SD = 74).
Mean looking times for each type of test sequences are presented in Figure 2 (left part). Since looking times were not normally distributed in the High TPs sequences condition as indicated by Shapiro–Wilk tests, $D(21) = .880, p = .015$, all data were log-transformed prior to analysis.

A paired sample $t$ test on these log-transformed looking times revealed that infants’ looking times differed between High and Low TPs sequences, $t(20) = 2.223, p = .038$, Cohen’s $d = .485$. On average, infants looked longer at Low than at High TPs sequences (see Figure 2). Fourteen of the infants (67%) showed this preference. In accordance with previous findings, the present experiment thus shows the infants’ ability to discriminate between highly familiar and less familiar visual sequences.

The average difference in log-transformed looking times between Low and High TPs sequences did not correlate with the length of the habituation phase, $r = -.403, p = .070$.

### 2.3 Discussion

In this first experiment, we replicated and extended previous findings that infants can discriminate between highly familiar and less familiar sequences and that they show a novelty preference in that context (Slone & Johnson, 2015). Infants were indeed faced with at least two additional levels of complexity in Experiment 1 than in Slone and Johnson’s (2015) experiments. First, the inter-shapes interval was made constant in our study so that infants had no cue to find pairs boundaries except the TPs. Slone and Johnson (2015), by contrast, used a 750-ms pause between any two test pairs of shapes. Second, we did not repeat any of our test trial whereas Slone and Johnson (2015) presented infants with three repetitions of two unique test trials. Hence, our
results show that infants can discriminate between highly familiar and less familiar sequences even in the absence of any cue boundaries and with no opportunity to learn the regularities during the test phase itself.

Nevertheless, these results do not allow determining with certainty whether infants discriminated between both types of pairs based on their (lack of) familiarity with the pairs, or based on the difference in the value of the transitional probability between pair elements (i.e., whether they considered low TPs pairs as unfamiliar or novel, or as familiar, but differing from high TPs pairs in terms of the value of TPs). Therefore, in a second experiment, we contrasted Low and Null TPs test pairs. The rationale is that if participants encode the value of the TPs between pair elements, they should consider Low TPs pairs as somehow familiar, at least relative to Null TPs pairs that they never encountered in the habituation phase. By contrast, if participants only encoded High TPs pairs as familiar in Experiment 1, namely if participants only chunked together elements co-occurring with a high probability, they would not discriminate Low from Null TPs pairs in Experiment 2 and would consider both as unfamiliar or novel.

3 | EXPERIMENT 2: LOW VERSUS NULL TPs

3.1 | Method

3.1.1 | Participants

Twenty-three full-term 7- to 9-month-old infants (11 females) composed the final sample (M age = 262 days, SD = 26, range = 210–297). Five additional infants were tested but excluded from the analyses due to fussiness (n = 2), or failure to meet the habituation criterion (n = 3). All infants had no known developmental difficulties. Recruitment procedures and thank-you gifts were the same as in Experiment 1.

3.1.2 | Apparatus and stimuli

Apparatus and stimuli were the same as in Experiment 1.

3.1.3 | Procedure

The procedure was the same as in Experiment 1 except that the two different types of test trials differing according to the TPs between shapes were Low and Null TPs (3 trials per condition, chosen randomly for each infant among all possible pairs in each condition). Low TPs sequences were the same as in Experiment 1. In the Null TPs sequences (Figure 1d), the pair consisted of two shapes appearing in the first position of two different pairs from the habituation phase, or of two shapes appearing in the second position of two different pairs from the habituation phase, so that TPs from one shape to another were always equal to 0. Test trials were presented pseudo-randomly so that a single type of test trial (Low or Null TPs) was not presented more than twice in a row.

---

4This sample size provides a power of .89, given a .05 significance criterion in a one-tailed test and an effect size of .62 (as reported by Slone & Johnson, 2015, in the older group of infants).
3.2 | Results

Infants were presented with 5–12 habituation trials (M = 7.964, SD = 1.988). Mean duration of the habituation phase was 148 s (SD = 78).

Mean looking times for each type of test sequences are presented in Figure 2 (middle part). Since looking times were not normally distributed in the Low and Null TPs sequences condition as indicated by Shapiro–Wilk tests, D(23) = .873 and .909, and p = .007 and .04, all data were log-transformed prior to analysis.

A paired sample t test on these log-transformed looking times revealed that infants’ looking times did not differ between Low and Null TPs sequences, t < 1 (see Figure 2). To ascertain that this null result denotes an absence of sensitivity to Low and Null TPs sequences (null hypothesis), we computed a Bayesian paired sample t test using JASP (2019) and obtained a BF of .251. Since values below 1/3 constitute evidence for the null hypothesis (Jeffreys, 1961), this analysis clearly supports that infants did not look differentially to Low and Null TPs sequences in our experiment.

The average (non-significant) difference in log-transformed looking times between Low and Null TPs sequences did not correlate with the length of the habituation phase, r = .150, p = .495.

3.3 | Discussion

In this experiment, infants did not discriminate between Low and Null TPs pairs. At first sight, this finding thus supports that infants did not encode the value of the TPs between pair elements. Taken together with findings from Experiment 1, these results suggest that infants only chunked together elements co-occurring with a high probability (i.e., High TPs pairs) and therefore considered Low and Null TPs pairs as unfamiliar even though they differ according to the strength of their TPs.

Further support for this possibility would nevertheless come from the examination of the infants’ ability to differentiate between these different types of pairs when presented in the same test phase. In Experiment 3, High, Low, and Null TPs pairs were thus presented in a within-subject design. Based on results from Experiments 1 and 2, we expected that infants would discriminate between High and Low TPs pairs, but not between Low and Null TPs pairs.

4 | EXPERIMENT 3: HIGH VERSUS LOW VERSUS NULL TPs

4.1 | Method

4.1.1 | Participants

Twenty-five full-term 7- to 9-month-old infants (20 females) composed the final sample (M age = 252 days, SD = 25, range = 217–288).\(^5\) Eleven additional infants were tested but excluded from the analyses due to fussiness (n = 8), or failure to meet the habituation criterion (n = 3). One other infant was excluded from the sample because her average log-transformed looking looking times were not normally distributed.

---

\(^5\)This sample size provides a power of .91, given a .05 significance criterion and an effect size of .62 (as reported by Slone & Johnson, 2015, in their older group of infants).
time in one condition was more than two standard deviations away from the average log-transformed looking time of all participants on test trials of the same condition (see below). All infants had no known developmental difficulties. Recruitment procedures and thank-you gifts were the same as in Experiments 1 and 2.

4.1.2 | Apparatus and stimuli

Apparatus and stimuli were the same as in Experiments 1 and 2.

4.1.3 | Procedure

The procedure was the same as in Experiments 1 and 2 except that there were three different types of test trials differing according to the TPs between shapes: High, Low, and Null TPs (2 trials per condition, chosen randomly for each infant among all possible pairs in each condition). The test phase was virtually divided into two blocks of three randomly presented trials, each of them corresponding to a single exemplar of each type of test trial. Mean looking times across each of the three conditions were computed.

4.2 | Results

Infants were presented with 5–12 habituation trials (M = 7.56, SD = 2.219). Mean duration of the habituation phase (i.e., the total looking time at the screen) was 186 s (SD = 93).

Mean looking times for each type of test sequences are presented in Figure 2 (right part). Since looking times were not normally distributed in the High and Null TPs sequences conditions as indicated by Shapiro–Wilk tests, $D(25) = .900$ and $.878$, both $p < .05$, all data were log-transformed prior to analysis.

A repeated measures analysis of variance (ANOVA) was applied on these log-transformed looking times, with type of test sequence as a within-subject factor (3 levels: High, Low and Null TPs). The effect of type of test sequence was significant, $F(2, 48) = 4.784$, $p = .013$, partial $\eta^2 = .166$.

Contrary to our predictions based on previous studies (Slone & Johnson, 2015) and on results from Experiment 1, the difference in looking times between High and Low TPs sequences was not significant, $t < 1$. To ascertain that this null result denotes an absence of sensitivity to High and Low TPs sequences (null hypothesis), we computed a Bayesian paired sample $t$ test using JASP Team (2019) and obtained a BF of .213. Since values below 1/3 constitute evidence for the null hypothesis (Jeffreys, 1961), this analysis clearly supports that infants did not look differentially to High and Low TPs sequences in our experiment. Also conflicting with our predictions based on results from Experiment 2, the difference in looking times for Low and Null sequences was significant, $t(24) = 3.062$, $p = .005$, Cohen’s $d = .612$. On average, infants looked longer at Low than at Null TPs sequences, supporting that, in the test context of Experiment 3, they discriminated between less familiar and unfamiliar sequences. This preference was exhibited by 16 of the infants (64%). Finally, as predicted, planned comparisons revealed that infants’ looking times differed between High and Null TPs sequences, $t(24) = 2.632$, $p = .015$, Cohen’s $d = .526$. Nevertheless, on average, infants looked longer at High than at Null TPs sequences (see Figure
2). Looking at individual data, we observed that this pattern of results was exhibited by 16 of the infants (64%).

The average difference in log-transformed looking times between High and Null TPs sequences did not correlate with the length of the habituation phase, $r = -0.241$, $p = .245$. The same was true regarding the average difference in log-transformed looking times between Low and Null TPs sequences, $r = -0.007$, $p = .974$.

Of note, a one-way ANOVA applied on mean durations of habituation phases showed no effect of Experiment (Experiment 1: $184 \pm 93$ s, Experiment 2: $148 \pm 78$ s, Experiment 3: $197 \pm 66$ s), $F(2, 66) = 2.26, p > .10$.

4.3 Discussion

Results from Experiment 3 contrast on three points with the predictions made on the basis of the results of Experiments 1 and 2: (1) No difference in looking times was observed between High and Low TPs sequences, (2) infants showed a clear discrimination of Low and Null TPs sequences (and of High and Null TPs sequences), and (3) infants looked longer at High and Low TPs pairs than at Null TPs pairs (i.e., we observed a familiarity instead of a novelty preference). These results therefore strongly support that (1) infants do extract the strength of the TPs, but that (2) expressing that knowledge would vary as a function of which and how many types of sequences are contrasted in the test phase. As a matter of fact, while the habituation phases were strictly identical in all three experiments, test phases were different: while two types of test sequences were used in Experiments 1 and 2 (High and Low, and Low and Null TPs pairs, respectively), three types were presented in Experiment 3 (High, Low and Null). Hence, the diversity of the test trials was higher in Experiment 3. This might have affected the global pattern of results.

Nevertheless, when starting the test phase, infants did not know in advance to which and how many different test sequences they would be exposed to. In particular, in Experiment 3, it was only from the third trial that infants had to deal with a more diverse, complex test phase (i.e., a third type of test sequence) than in Experiments 1 and 2. We therefore reasoned that the early test arrangement (i.e., the arrangement made by the two first test trials) might have determined the looking patterns toward the three types of test sequences in Experiment 3 and be responsible for the unexpected pattern of results.

If this assumption is correct, then (1) infants in Experiment 3 who started the test phase with a similar early arrangement as (most of) the infants in Experiment 1 (i.e., with a High and a Low test pairs, in any order, i.e., High then Low or Low then High) should show a novelty preference as infants in Experiment 1, and (2) infants in Experiment 3 who started the test phase with a similar early arrangement as (most of) the infants in Experiment 2 (i.e., with a Low and a Null

---

6In Exp. 1, 15 infants started with a High and a Low test trials, while three started with a succession of two High test trials and three with a succession of two Low test trials. Importantly, analyses that considered only the infants from Experiment 1 who started with a High and a Low test trials revealed the same novelty effect as considering the whole sample, $F(1, 14) = 5.628, p = .033$, partial $\eta^2 = .287$.

7In Exp. 2, 15 infants started with a Low and a Null test trials, while five started with a succession of two Low test trials and three with a succession of two Null test trials. Importantly, considering only the infants from Experiment 2 who started with a Low and a Null test trials revealed the same null effect as considering the whole sample, $F < 1$, BF = .293).
test pairs, in any order) should show similar interest for Low and Null test pairs as infants in Experiment 2. In other words, the patterns of looking times in Experiment 3 should vary as a function of the early test arrangement.

4.3.1 Impact of the early test arrangement on infants’ looking times

To examine this possibility, we re-ran the previous analysis and included a new factor that defined for each participant what were his/her first two trials. Given that the test phase was virtually divided in two blocks (see Method), the new variable “First two trials” had three possible levels: High-Low (n = 10), Low-Null (n = 7), and High-Null (n = 8).

As expected based on the previous analysis, the effect of type of test sequence was significant, $F(2, 44) = 6.729, p = .003$, partial $\eta^2 = .234$. More interestingly, the interaction between type of test sequence and first two trials was also significant, $F(4, 44) = 4.269, p = .005$, partial $\eta^2 = .280$ (see Figure 3). The main effect of first two trials was not significant, $F(2, 22) = 1.163, p = .313$, partial $\eta^2 = .096$.

Simple main effects analyses revealed that the effect of type of test sequence was significant in the High-Low and in the Low-Null groups, $F(2, 18) = 5.042, p = .018$, partial $\eta^2 = .359$ and $F(2, 12) = 5.058, p = .026$, but not in the High-Null group, $F(2, 14) = 2.863, p = .091$.

In the High-Low group, infants looked longer at Low than High test trials, replicating the novelty effect found in Experiment 1, $t(9) = 2.664, p = .026$. They also looked longer at Low than at Null test trials, $t(9) = 2.334, p = .044$. Looking times to High and Null test trials did not differ significantly, $t < 1$, BF = .316.

**Figure 3** Mean looking times for each type of test sequences in Experiment 3, as a function of the early test arrangement defined by the first two test trials, that is, High and Low, High and Null, or Low and Null test trials (in any order). Error bars represent 95% confidence intervals.
In the Low-Null group, while infants did not look differently at Low and Null test trials, replicating the null effect observed in Experiment 2, \( t < 1, BF = .216 \), they looked significantly longer at High than at Low and Null test trials, \( t(6) = 2.484 \) and \( 2.527, p = .048 \) and .045, respectively.

These results clearly demonstrate that the early test arrangement made up of the first two test trials\(^8\) determines the infants’ interest for visual sequences at test. We discuss this point further in the next section.

5 | GENERAL DISCUSSION

Several recent studies have convincingly demonstrated the infants’ ability to detect sequential regularities between visual objects (Kirkham et al., 2002; Slone & Johnson, 2015; Tummeltshammer et al., 2016). Nevertheless, the extent to which infants acquire representations of the relations between these objects that include the value of the transitional probabilities (TPs) between them, and are therefore able to make fine-grained distinctions between sequences involving various levels of TPs, remains an open question. The present study aimed at addressing this issue by examining 8-month-old infants’ ability to discriminate between high and low TPs sequences, namely highly familiar and supposedly less familiar sequences (as in previous studies), but also between these two types of sequences and completely unfamiliar ones (i.e., null TPs sequences).

Results showed that 8-month-olds could extract subtle statistical information from a visual stream. As a matter of fact, infants not only differentiated between highly and less familiar pairs of shapes (Experiment 1, and High-Low group in Experiment 3), but they also discriminated between completely unfamiliar sequences and familiar ones, be they either highly or less familiar (Experiment 3). Hence, rather than having merely extracted and learned the most salient regularities within the visual stream they were exposed to, infants were also sensitive to subtler regularities and subsequently able to discriminate between sequences based on the strength of these regularities (see Tummeltshammer & Kirkham, 2013, for similar findings in the spatiotemporal domain). These findings thus constitute a strong argument in favor of a statistical learning mechanism by which infants became sensitive to the actual values of the statistical regularities between shapes when exposed to a stream of visual stimuli. Possibly, that mechanism could be a step toward the building of chunks (Slone & Johnson, 2018).

Importantly, the expression of this statistical knowledge\(^9\) would be sensitive to the type of test sequences and their arrangement at test. As a matter of fact, although in three experiments learning outcomes were assessed in comparable groups of infants following strictly identical learning phases (hence at the end of which infants should have acquired similar

---

\(^8\)Of note, considering in the analysis only the very first trial, namely whether the test phase started with a High, Low, or Null TPs sequence \((n = 11, 9, \text{ and } 5, \text{ respectively}), \) neither the main effect of First trial nor the interaction between Type of test sequence and First trial were significant, \( F(2, 22) = 2.353, p = .119, \) partial \( \eta^2 = .176 \) and \( F(4, 44) = 1.468, p = .228, \) partial \( \eta^2 = .118. \)

\(^9\)By « expression of knowledge », we do not mean to suggest that children have developed intentionally accessible propositional knowledge but that they have, at a minimum, developed a sensitivity to the statistical regularities of the sequence of shapes that will effectively influence their behavior at test.
knowledge of the regularities between stimuli), we observed that the infants’ interest toward the different types of sequences varied as a function of the experiment, and more specifically as a function of the first two test trials. While infants looked longer to Low than High TPs sequences when the test phase started with a High and a Low TPs test sequences (Experiments 1 and 3), they looked less to Low than High TPs sequences when the test phase started with a Low and a Null TPs test sequences (Experiment 3). Similarly, although infants showed similar interest for Low and Null TPs sequences when the test phase started with a Low and a Null TPs test sequences (Experiments 2 and 3), they looked longer at Low than at Null TPs sequences when the test phase started with a High and a Low TPs test sequences (Experiment 3). These results strongly suggest that the expression of infants’ discrimination abilities of visual sequences based on the strength of the between-shapes associations is conditioned upon the early arrangement of the test phase, operationalized in the present study as the two first items. These results also highlight that the first two items determine the pattern of looking times to these two types of sequences, no matter if a third type of sequence is presented afterward.

But what in the early trial arrangement determines the infants’ looking behavior toward sequences involving strongest or lowest levels of TPs? We propose that the predictability of the first test items—namely whether these items were either expected or unexpected based on what infants have been habituated to in the learning phase—conditioned the infants’ differential interest for test sequences differing by their TPs. Numerous studies have indeed shown that infants can form learning-induced expectations (Emberson et al., 2015). Yet, in our study, the test phase immediately followed the habituation phase in which High and Low TPs were presented alternately. Hence, starting the test phase with a High and a Low TPs sequences (as most infants in Experiment 1, and about one third of them in Experiment 3) was much less surprising than having a Low and a Null TPs sequences as first test trials (as in most infants in Experiment 2, and about one third of them in Experiment 3).

Our data show that when the test phase starts with non-surprising test sequences, infants allocate their attentional resources to the least familiar stimuli. Remarkably, when completely unfamiliar sequences are presented afterward (as for the High-Low group in Experiment 3), infants do not show special interest for these novel sequences: they indeed looked less to these sequences than to Low TPs sequences. This pattern of results closely mimics the so-called Goldilocks effect (Kidd et al., 2012), according to which infants prefer stimuli that are neither too simple nor too complex based on their current representation of the world, namely on what they have been exposed to beforehand. Infants therefore attend more to moderately predictable events, the Low TPs sequences, and look away from events that are either especially predictable, the High TPs sequences, or especially surprising, the Null TPs sequences.

In unpredictable test arrangements (i.e., with initial test trials that strongly depart from the habituation phase), infants were equally interested to less and unfamiliar trials, probably because they constitute a surprising, unexpected arrangement. When highly familiar sequences are added afterward (as for the Low-Null group in Experiment 3), infants devote extra attention to these sequences, showcasing their statistical knowledge.

A novelty preference was observed when the test phase started in a predictable manner, and a familiarity preference was observed when the first test trials were at odds with infants’ expectations. Interestingly, such a pattern of results fits well with Hunter and Ames’ (1988) model. This model indeed proposes that, all else being equal, easy tasks would lead infants to show a novelty preference, while more complex tasks would lead to a familiarity preference. In a recent study, Emberson et al. (2019) indeed argued that the familiarity preference they observed arose from the
use of complex stimuli (faces), which rendered the task more difficult for the infants. Arguably, the infants’ task in our study was more complex when the test phase started with unexpected versus predictable events. Indeed, in the former case, the TPs of the first test trials contrast with the learning material, and infants would need to adjust the predictions they made based on the habituation phase.

Obviously, although these interpretations are appealing, future studies should be specifically dedicated at examining the impact of the order of the test trials and of their predictability on infants’ interest to visual sequential events differing by the strength of their TPs. In particular, future studies should take special care in controlling the order of presentation of the (different types of) test stimuli. These indeed constitute the test context for every ensuing test item, so that their order of appearance could affect the distribution of infants’ looking times.

To sum up, our results indicate that 8-month-old infants can learn fine-grained statistical regularities between visual shapes occurring in succession. Importantly, the expression of their statistical knowledge would depend on the arrangement of the test trials and more specifically on the predictability of the first test items, hence supporting that infants’ SL is a flexible mechanism (Tummeltshammer et al., 2016). Further studies are clearly needed to confirm the influence of the composition and arrangement of the post-learning phase on infants’ discrimination abilities and preference, but the present results yet provide clear evidence that (1) the absence of preference cannot be decidedly considered as reflecting the infants’ inability to discriminate between sequences and that (2) researchers should carefully consider the nature and arrangement of the different types of test. The question of whether and how infants use that subtle statistical knowledge not only to discriminate between visual series, but also to optimize their anticipation and planning behaviors, should be addressed in further studies.

ACKNOWLEDGMENTS
This work was supported by a FRS-FNRS grant (F.4524.10). The first author (J.B.) was a Post-Doctoral Researcher of the Fonds de la Recherche Scientifique-FNRS. The second author (E.S.A.) was a Research Fellow of the Fonds de la Recherche Scientifique-FNRS. The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID
Julie Bertels "https://orcid.org/0000-0002-0859-825X
Hermann Bulf "https://orcid.org/0000-0003-4121-1341

REFERENCES


JASP Team. (2019). JASP (Version 0.10.2) [Computer software].


