

Distributional Learning of Categories in Infants: From Phonemes to Actions

Honors Thesis

Presented to the College of Arts and Sciences,

Cornell University

in Partial Fulfillment of the Requirements for the

Biological Sciences Honors Program

by

Lucas Chang

May 2013

Supervisor: Michael Goldstein

Abstract

Infants can perceive categorical differences in speech sounds based on the distribution of features in a set of sounds they are briefly exposed to (Maye et al, 2001). The present work proposes a mechanism for this category learning and implements it in a dynamic field model, reproducing the effect in which bimodally distributed features led to category discrimination and unimodally distributed features led to non-discrimination. The categorical effect is therefore explainable as an emergent property of the dynamics of a domain-general perceptual-cognitive learning system. Then, an experiment was designed to investigate whether the same mechanism could be used domain-generally to categorize actions. 10.5-12-month infants viewed a set of animated actions that differed in a feature that had either a unimodal or bimodal distribution. Looking time and rate of looking away were measured to determine whether the distributional information in the input modulated attention and whether infants learned to perceive the actions categorically.

Introduction

How do infants identify actions occurring around them? Infants are constantly confronted with a multitude of people and objects moving and interacting in various ways, seemingly without obvious cues to when actions begin and end, or which actions fall into the same category. For example, when an adult reaches for a toy and gives it to the infant, what portion of the action corresponds to reaching and what portion to giving? When Mom's arm extends toward the toy, is she reaching for it or pointing to it? Identifying such actions is a crucial prerequisite for tasks such as verb learning and understanding of others' goals. Current theories often take a top-down approach in which knowledge that others are intentional agents drives the development of action perception. However, these accounts undervalue the role of basic perceptual and statistical learning mechanisms. The current research investigates whether a statistical learning mechanism infants use to categorize speech sounds can be extended to categorizing actions.

Infants have powerful mechanisms for extracting structure from continuous perceptual input, and many of these mechanisms have been investigated in detail in the domain of language

acquisition. Human language, especially the speech people use around infants, has well-documented predictable structure on multiple levels [1, 2]. Moreover, infants are sensitive to many of these regularities [3]. In recent decades, extensive research demonstrated that prelinguistic infants use multiple mechanisms to learn regularities in structured linguistic input [4]. These statistical learning mechanisms have been shown to be domain-general by replicating language-learning experiments using visual stimuli or nonlinguistic sounds such as tones or animal sounds [5]. This suggests that infant statistical learning is not merely an adaptation for language acquisition, but rather may be involved in learning meaningful regularities from structured experience in several domains, including the perception of action.

Concurrently with segmenting an action sequence into its components, infants must categorize the resulting segments before they can be useful for verb learning or understanding others' intentions. They must, for instance, reliably distinguish a reach from a point in order to map these actions onto words or predict whether someone intends to grab an object or to communicate. Categorization of actions has been studied with a framework grounded in the semantics of motion verbs in language. Although the characteristics of a motion can be described in several ways, natural languages have a strong tendency to encode them in verbs according to *path* and/or *manner*. For example, English verbs encode a distinction in manner between *walk* and *run* in "The cow walked/ran into the barn," and a distinction in path between *enter* and *leave* in "The cow entered/left the barn." After exposure to a set of actions with one of manner or path constant and the other variable, infants under a year can discriminate a novel action that preserves the invariant feature from one that is novel in both features [6].

In an alternative, perceptually grounded framework inspired by research on face processing, infants use "featural" and "configural" information to categorize actions [7]. Featural

information consists of local details (e.g., scratching vs brushing the shoulder), while configural information consists of more global properties (e.g., a straight vs arcing path toward an object). Infants more easily distinguished between actions with featural changes, even when configural changes were much greater in magnitude.

What accounts for biases for local detail in categorizing actions? Infants could have an evolved bias to attend to featural information, which would be adaptive because such information is more useful for mapping actions to words and deciphering other people's intentions. Alternatively, infants could learn to attend selectively to certain aspects of actions based on prior experience, as in perceptual narrowing, in which young infants initially discriminate between a wide variety of different stimuli, then gradually lose the ability to discriminate between stimuli with which they have less experience. Perceptual narrowing occurs most rapidly during the second half of the first year, when infants lose the ability to distinguish between stimuli, such as monkey faces or phonemes from other languages, that are not relevant to their developmental context [8]. As infants lose the ability to discriminate some stimuli, they also begin to perceive discrete categories, discarding within-category variation as irrelevant. Thus, perceptual experience guides infants to the most useful sources of information.

Previous work, however, has left unspecified the nature of infants' representations for actions. Do infants have conceptual knowledge about actions, or do they perceive actions using lower-level features? In order to investigate this issue, the current study introduces actions that vary systematically along a perceptual variable.

Distributional Learning

Maye et al. [9] demonstrated a particular unsupervised statistical-learning mechanism that can account for perceptual narrowing of phonemic contrasts. Infants heard stimuli along a

continuum between two phonemes, [d] and [t], which varied along two parameters, voice onset time and formant structure. One group heard a set of sounds with the features in a unimodal distribution, while the other group heard them in a bimodal distribution, indicating the presence of a category distinction (Fig.1). Only the bimodal group subsequently distinguished between stimuli from opposite sides of the continuum. Thus, when the distribution of inputs along a particular dimension reflects a relevant distinction, infants learn to attend selectively to that dimension. In addition, a recent study showed that the same type of learning can account for both loss of sensitivity to features that occur in a unimodal distribution and gain of sensitivity to features that occur in a bimodal distribution [10]. If this learning mechanism generalizes to the domain of action perception, then quantitative features of actions that occur in bimodal distributions should result in behavior consistent with categorization, while unimodal distributions should not. This can potentially account for the finding of increased sensitivity to featural information; local details may be more salient because they tend to fall in roughly discrete categories, while global properties may tend to be more uniformly distributed.

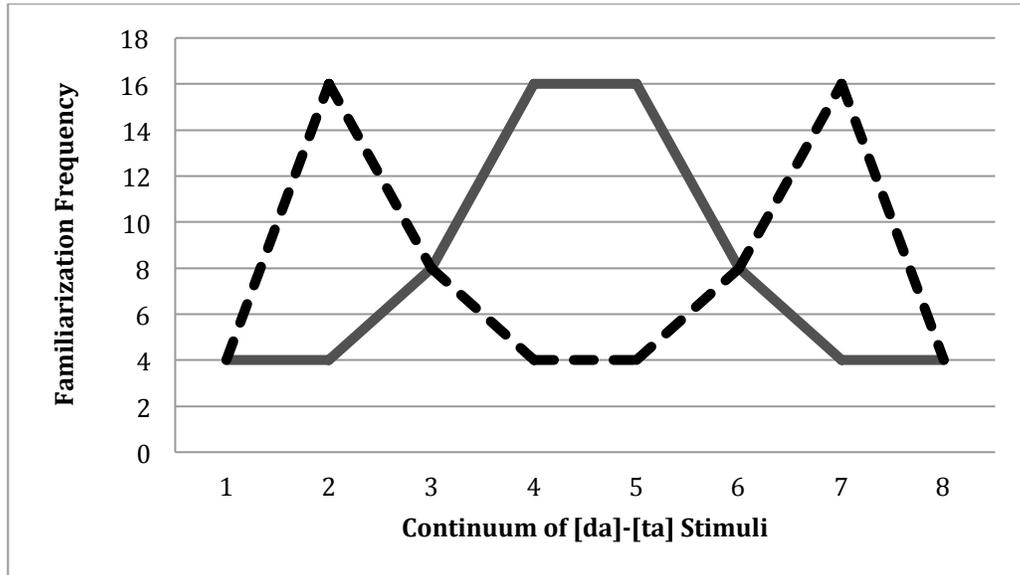


Figure 1: Reproduced from Maye et al. 2002. Bimodal vs. Unimodal distributions of stimuli used for familiarization. The distribution for the Bimodal group is shown by the dashed line, and the Unimodal group by the solid line.

Categorization

Although categorization in infants has been extensively studied, research on the mechanisms of category formation has been scarce [11]. One approach has used autoencoder networks to simulate categorization. These networks work by receiving the input in one layer and reconstructing it in an output layer, after passing through a hidden layer, which crucially has fewer units than the input layer. The hidden layer therefore cannot represent the input with full detail, but learns to encode a more compact representation. Mareschal and French [12] simulated data from Younger's 1985 study [13], in which infants saw a set of line drawings of animals, whose features (leg length, ear separation, etc.) were correlated in one group, indicating a category distinction, and uncorrelated in another, indicating the lack of a distinction. The network reproduced infant looking-time data and the internal representation in the hidden layer

showed category-like behavior: when features were correlated, the internal representation formed two clusters when activation in the three hidden units was plotted on three axes, while in the uncorrelated condition the representations were scattered throughout the representation space. What is described on the surface as “categorization” may be conceptualized as a distortion of the space in which items are represented, such that the distance between within-category items decreases and that among categories increases. This account is supported by evidence that, in adults, experience with categorically structured visual stimuli results in better discrimination between categories than within a category [14]. A similar distortion-based model comes naturally out of dynamic field models, discussed below.

Dynamic Field Theory

Dynamic field theory (DFT) is a framework for modeling neural computation that incorporates principles of cortical networks [15, 16]. The architecture involves layered fields of activation laid out along continuous spatial or feature dimensions. Input from the environment drives activation in a subset of these layers, while the intrinsic dynamics of the fields cause activation to evolve over time. Locally excitatory connections within a layer and laterally inhibitory interactions between layers allow activation to become self-sustaining, and persist in the absence of input. Layers with slower timescales of activation and decay provide memory that allows the system to store representations of previously encountered input. Excitatory and inhibitory connections between layers allow memory storage, influences of memory on processing, and other computations. Variations of this structure have been used to simulate many empirical findings, including but not limited to visually guided reaching, spatial cognition, habituation, and categorization [17, 18, 19]. Simulating distributional learning with a domain-general model links the proposed learning mechanism with cognitive dynamics more broadly.

Johnson, Spencer and Schöner [20] used DFT to simulate biases in spatial recall as well as recall in a nonspatial dimension, color. After experience with certain colors, the model accumulated a long-term memory trace for those colors. Then, while holding a color in short-term memory, the short- and long-term representations interact such that the remembered color shifts, as observed in behavioral studies. A similar mechanism underlies the simulation we present here, but we use a much more varied input set to lay the long-term memory trace, and we interpret directional biases induced by the shifting of internal representations as a global distortion of the representation space in a way that produces categorical behavior.

Habituation

The current study involves familiarizing infants with a corpus of animated actions. During exposure, infants are acquiring representations of these actions. *Habituation* occurs when repeated exposures to the same stimulus result in diminishing response [21], and *dishabituation* refers to the recovery in response that occurs to a new stimulus if it is successfully discriminated from the old one. As infants habituate to the display, therefore, the pattern of looking times to each animation can provide information about the representations they are forming. Infants are active information gatherers, and the control of visual attention is correspondingly complex and incompletely understood. Therefore, a secondary goal of the current study is to investigate attention to the stimulus as a function of both time and the distribution of previously seen stimuli. Infants allocate less attention to stimuli that are too simple (e.g. already understood or uninteresting) or too complex (e.g. unlearnable or random) [22, 23]. Although the precise mechanism of this effect likely differs among contexts and modalities, a unifying principle is that attention and the formation of representations (i.e. learning) interact bidirectionally and dynamically. Such “smart” allocation of attention derives at least in part from habituation.

However, while habituation-based paradigms are used extensively in developmental psychology, there are few studies investigating the dynamics of interactions between attention and learning.

Here I present a DFT model that replicates distributional learning of categories as in Maye et al. Then, we exposed 10.5-12-month infants to a set of animated actions that differed along a quantitative, linear dimension. Half of infants observed a unimodal distribution and the other half observed a bimodal distribution. We recorded their looking times and frequency of looks away from the stimulus during familiarization, and during a test phase we asked whether their looking behavior was sensitive to changes in the stimulus. We hypothesized that infants would look more overall to the bimodal distribution during familiarization because it contained more learnable structure, which would engage their attention. Further, we hypothesized that infants in the Bimodal condition would discriminate cross-category changes in motion in the test phase, while infants in the Unimodal condition would not.

Study 1: Simulation

We used a DFT model was used with 4 fields (Fig. 2). to simulate distributional learning of speech-sound categories. Previously, McMurray et al [24] modeled distributional category learning as optimization of a mixture of Gaussians to fit the observed distribution. Such a model can account for distributional category learning, but is specific to this task and does not address its implementation in a distributed, dynamic system.

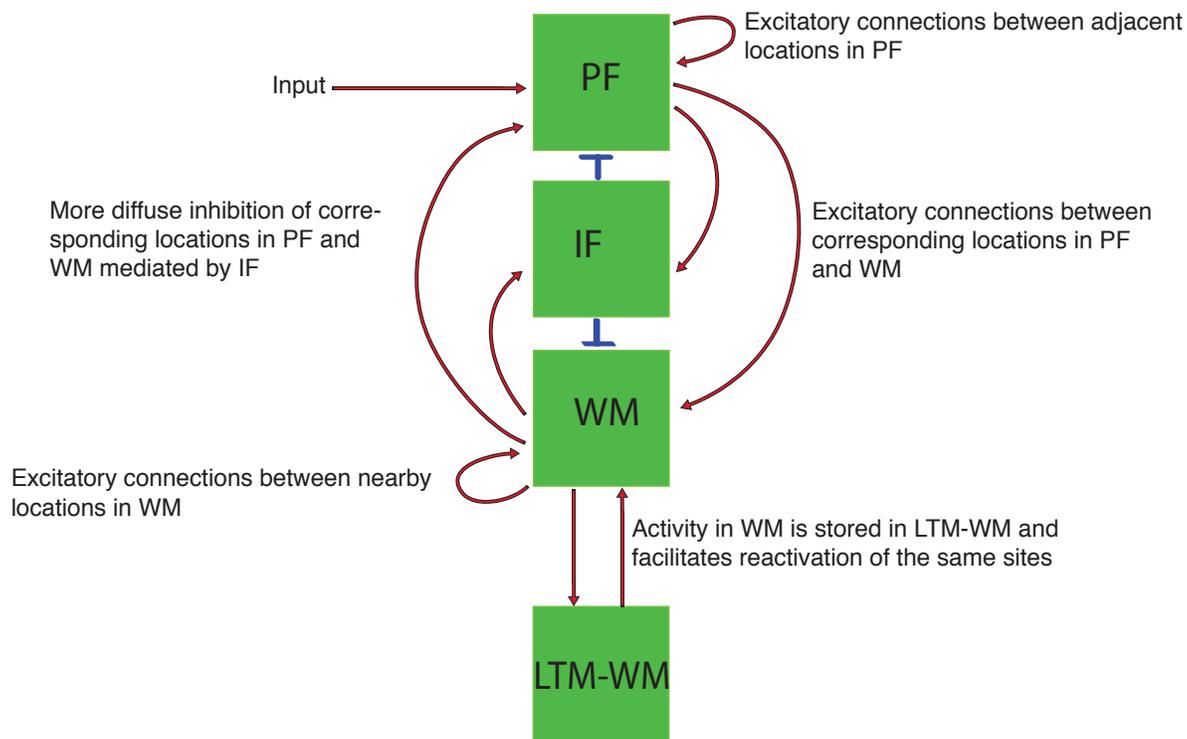


Figure 2: Schematic of model. Each field (Perceptual field PF, Inhibitory field IF, Working memory WM, Long-term memory store of working memory LTM-WM) consists of a one-dimensional array of 181 neurons, with connections within and between fields as shown and described in the text. Input was in the form of activity added directly to neurons in PF.

The model contains four fields, each of which consists of an array of 181 neurons. Each neuron is mapped to a point in the one-dimensional continuum of speech sounds, so that neurons closer to 1 represent sounds closer to [da] and neurons closer to 181 represent sounds closer to [ta]. Each neuron computes its continuous activity level at each time step. The change in activity of a neuron at each time step is the sum of terms representing decay towards a resting (inactive) level, input from other neurons and/or stimuli, and random noise (See Appendix for equations). If a neuron outputs to another neuron, the other neuron receives input according to a sigmoid function of the first neuron's activity level.

Neurons in the perceptual field (PF) receive direct input from the stimulus. They also receive excitatory connections from other PF neurons and from neurons in the working memory field (WM), and receive inhibitory connections from neurons in the inhibitory field (IF). The strength of connection between two neurons is a Gaussian function of the distance between them in the one-dimensional array, so that local connections are stronger than distant connections; the variance and maximum height of this function is different for each pair of connected fields. Neurons in WM receive excitatory connections from other WM neurons and from neurons in PF, and receive inhibitory connections from neurons in IF. Neurons in IF receive excitatory connections from neurons in PF and from neurons in WM.

The variance of inhibitory connections is greater than that of excitatory connections, thus adjacent neurons within PF and WM are excitatory while more distant neurons, though they have weak excitatory connections, are overall inhibitory (mediated by IF). This results in the formation of stable peaks of activity: when a neuron becomes active, reciprocal excitatory connections with nearby neurons result in stable high activity in a few nearby neurons, which does not spread to the rest of the field. The weights are set so that an activity peak in PF leads to the formation of an activity peak in WM; once a peak forms in WM, inhibition is strong enough to abolish a peak in PF if it is in the same location. Thus, activity in PF is transient, while activity in WM can be sustained over time, allowing it to function as working memory.

To allow for long-term learning effects, WM is associated with a long-term memory layer (LTM-WM). LTM-WM has slow dynamics; its time constants of rise and decay in activity are much slower than the other fields, so activity persists over the course of the whole simulation. Whenever activity in WM is above a threshold, LTM-WM receives input in the corresponding location. Thus, it remembers what locations have been active in WM in the past. Later, when

activity peaks form in WM, they are facilitated by the presence of activity in LTM-WM at the corresponding locations. Further, if activity in LTM-WM is stronger on one side of the WM peak than the other, then the asymmetric input causes the WM peak to drift in the direction of higher LTM-WM activation. This introduces a history-dependent bias that can account for distributional learning.

The model was used to simulate the distributional learning study in Maye et al. (2002). In that experiment, the stimuli were synthesized speech sounds that differed along a one-dimensional continuum defined by voice onset time and formant structure, varying between [da] and [ta]. In the simulation, we assume the perceptual system extracts these features and model the input as activation directly passed to the perceptual field. Stimuli were represented as Gaussian inputs to the perceptual field, with constant variance and mean varying in proportion to the acoustic features. The model was given 64 familiarization trials, each of which contained one input stimulus, taken in random order from the distributions in Fig.1.

At the start of each trial, the perceptual field, inhibitory field, and working memory field were set to resting values, while the long-term memory field accumulated a memory trace across trials. After the start of the trial, stimulus was introduced and the activity in each field was allowed to evolve.

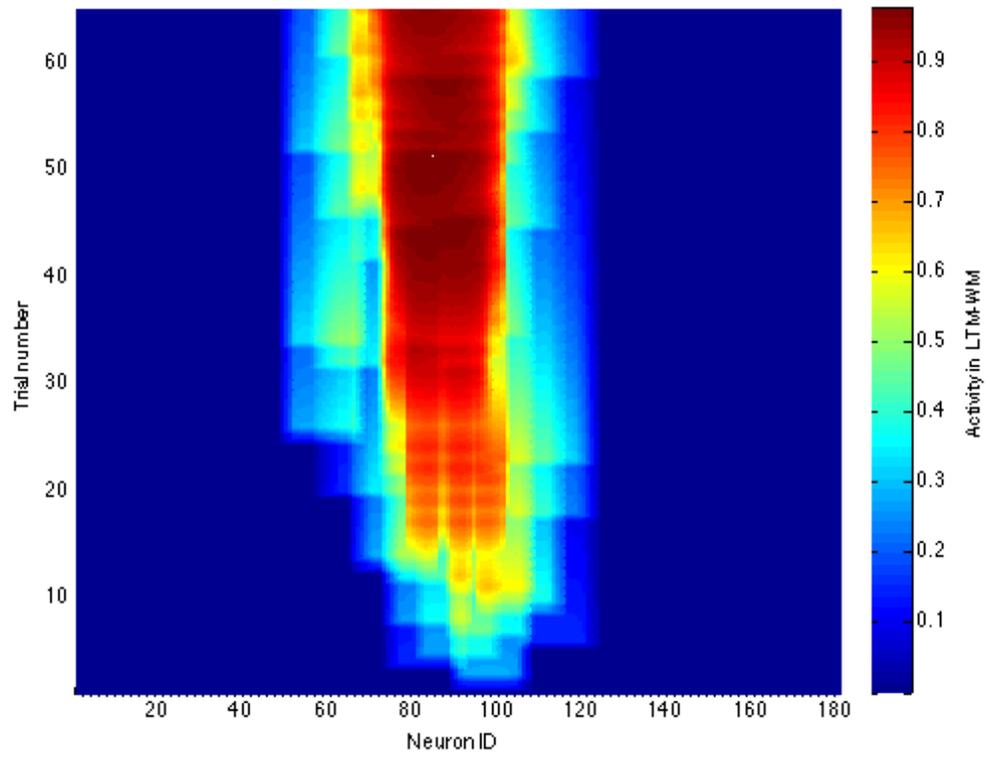
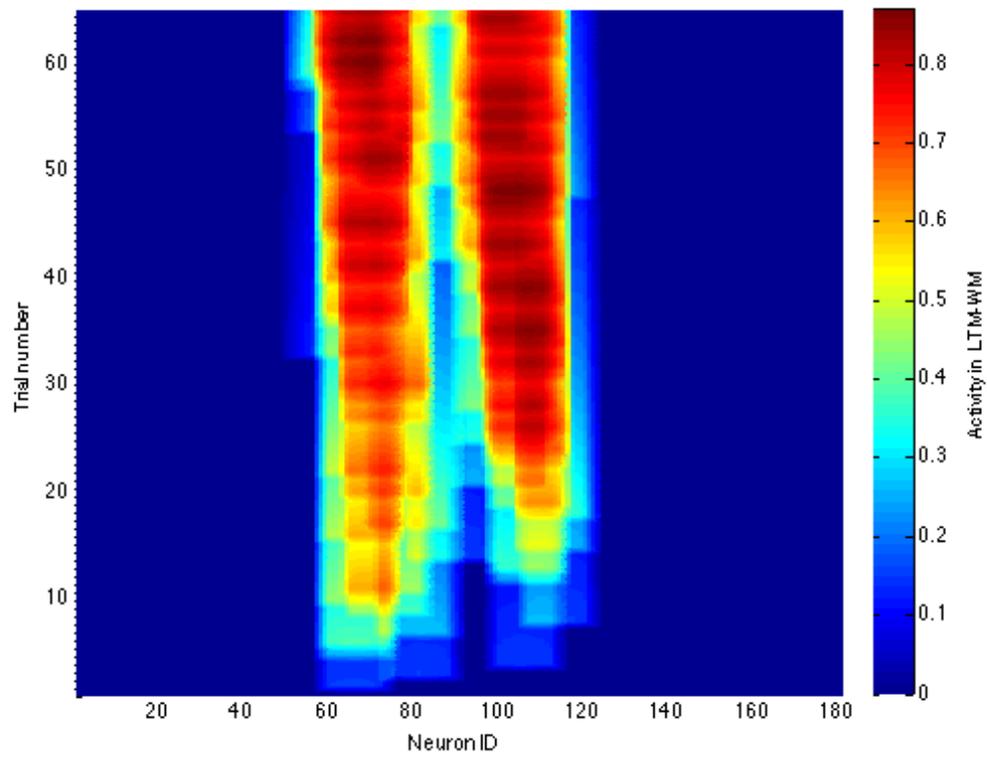
After the 64 familiarization trials, the “test” trial consisted of either two identical presentations of stimulus 3, or a presentation of stimulus 3 followed by a presentation of stimulus 6.. If the network learned to distinguish the two stimuli, these two test trials should induce a different pattern of activity; if it learned to ignore the difference, they should produce the same pattern of activity. In the model, a “same” response occurs when the perceptual field has no activation, because its activation is inhibited by the corresponding location being active in

working memory. A “different” response occurs when the perceptual field is activated, because any positions held in working memory are not close enough to inhibit the region where input occurs.

Because the distributions in Fig. 1 were chosen to equalize the range and the frequency of the test items, they could not be controlled for variance. The bimodal distribution had a greater variance than the unimodal distribution. Therefore, to control for the possibility that greater variance, rather than bimodality, in the input distribution drives discrimination, a Variance-Control condition was included in which the input distribution was identical to the unimodal distribution in shape, but stretched by a factor of 1.323 so that its variance was equal to that of the bimodal distribution.

Results

The LTM-WM field was able to represent a good approximation of the input distribution. The unimodal or bimodal activation in this field by the end of the familiarization period reflects the most frequently activated locations in WM over the course of many stimulus presentations (Fig. 3).

A**B**

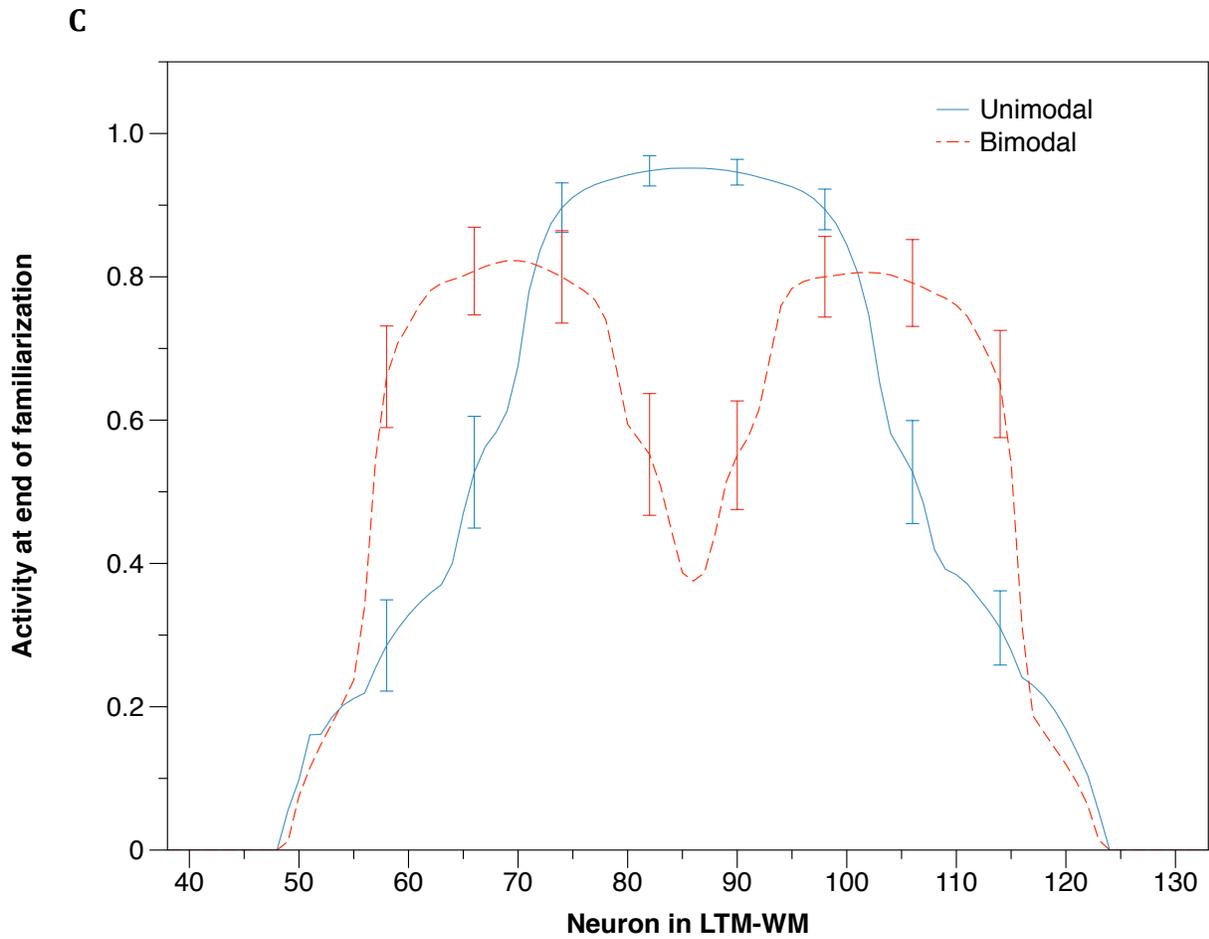


Figure 3: LTM-WM field builds up a representation of the input distribution over time. Representative time course of activity in LTM-WM, A: unimodal condition. B: bimodal condition. C: Mean (\pm SD) activity distribution in LTM-WM at the end of familiarization, data from 20 runs of simulation. Solid line: unimodal condition, Dashed line: bimodal condition

In Same test trials, the identical input was presented twice (Fig. 4.1). The remembered distributions (Fig. 4.5) cause the internal representation in WM and IF to drift towards the peaks (Fig. 4.3, 4.4) but in both cases the activity in IF still overlaps the input location, so the PF is inhibited and a second peak of activity is not formed at the second stimulus presentation (Fig. 4.2). Thus, in both conditions the model gives a Same response.

Unimodal condition

Bimodal condition

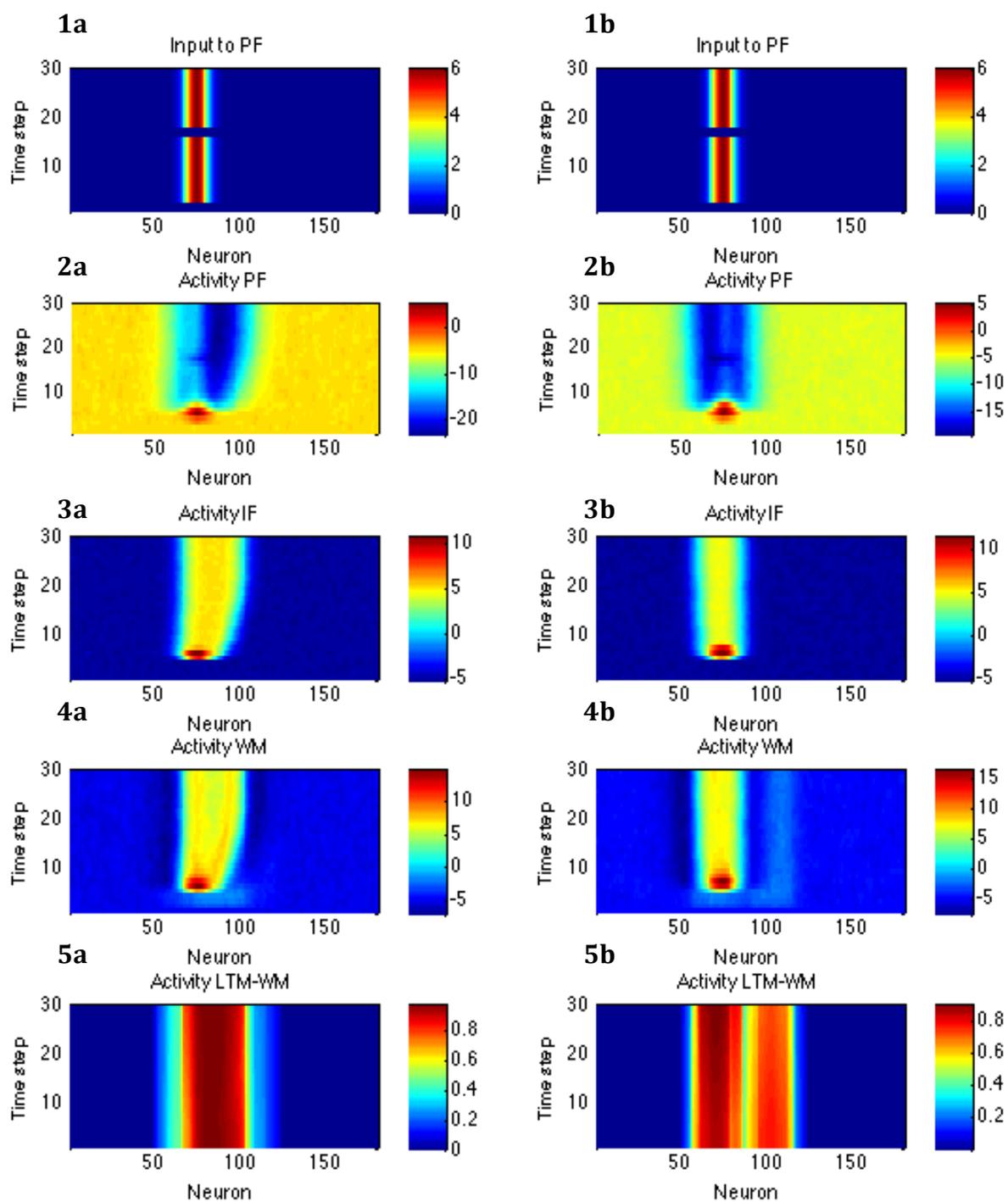


Figure 4: Activity during Same test trials, with position in the field on the horizontal axis and time on the vertical axis. Left: Unimodal condition Right: Bimodal condition. (1) Stimulus, (2) PF, (3) IF, (4) WM, (5) LTM-WM

In Different test trials, the two test inputs were on opposite sides of the overall mean (Fig. 5.1). The remembered distributions (Fig. 5.5) cause the internal representation in WM and IF to drift towards the peaks (Fig. 5.3, 5.4). In the unimodal condition, the drift is toward the single peak in the center, increasing the overlap with the location of the second stimulus, so the PF is inhibited and a second peak of activity is not formed at the second stimulus presentation (Fig. 5.2a). In contrast, in the bimodal condition the drift is away from the center towards one of the remembered peaks, decreasing the overlap with the location of the second stimulus. Therefore the PF is not inhibited, and a second peak of activity forms at the new stimulus presentation (Fig. 5.2b). Thus, the Bimodal condition gives a Different response while the unimodal condition gives a Same response. The Variance-Control condition produced the same pattern as the Unimodal condition, suggesting that the bimodal versus unimodal shape of the distributions, not variance, drives the differences in discrimination (Fig. 6), though this needs to be tested in infants.

Unimodal condition

Bimodal condition

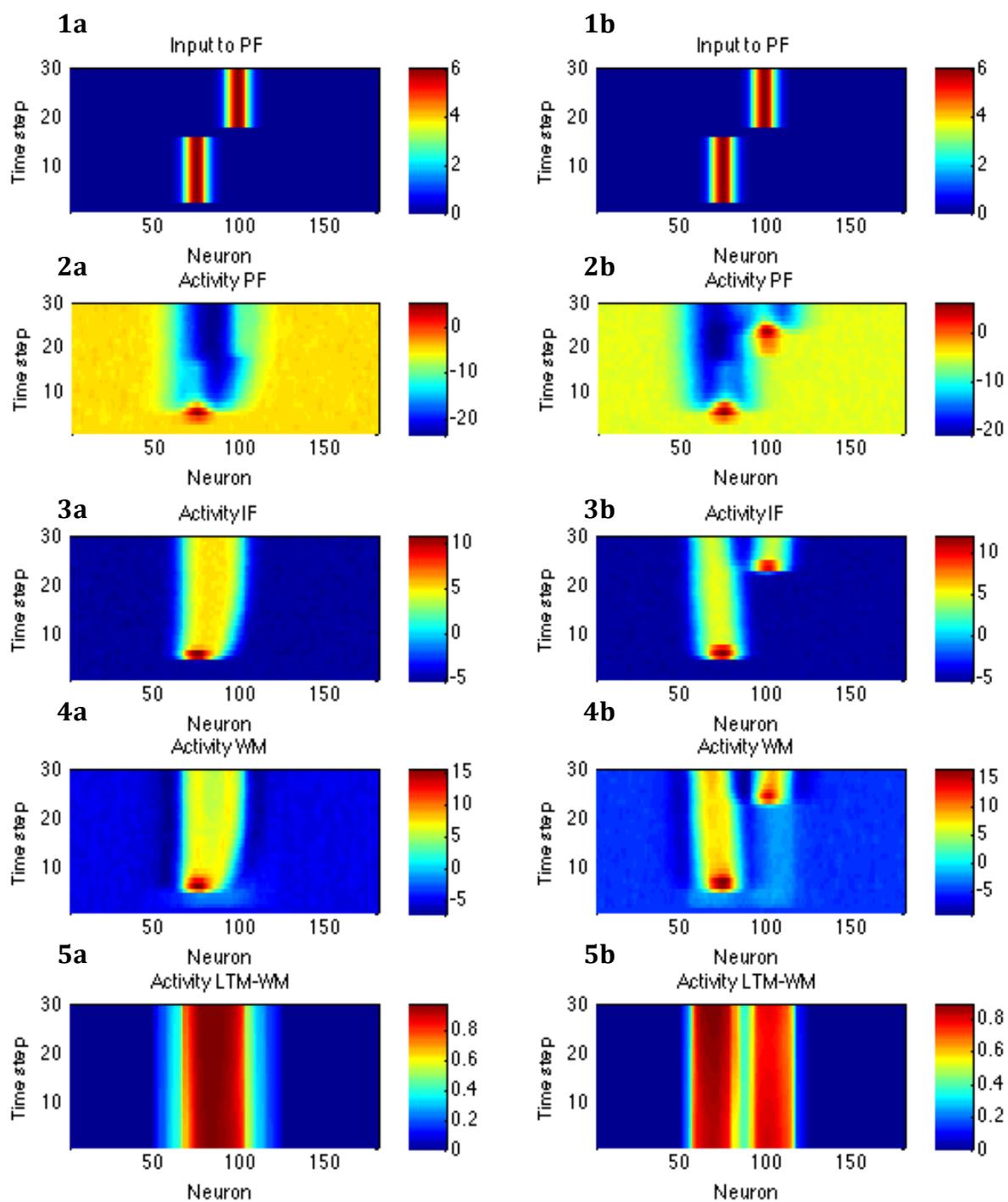


Figure 5: Activity during Different test trials, with position in feature space on the horizontal axis and time on the vertical axis. Left: Unimodal condition Right: Bimodal condition
1) Stimulus 2) PF 3) IF 4) WM 5) LTM-WM

Variance-Control condition

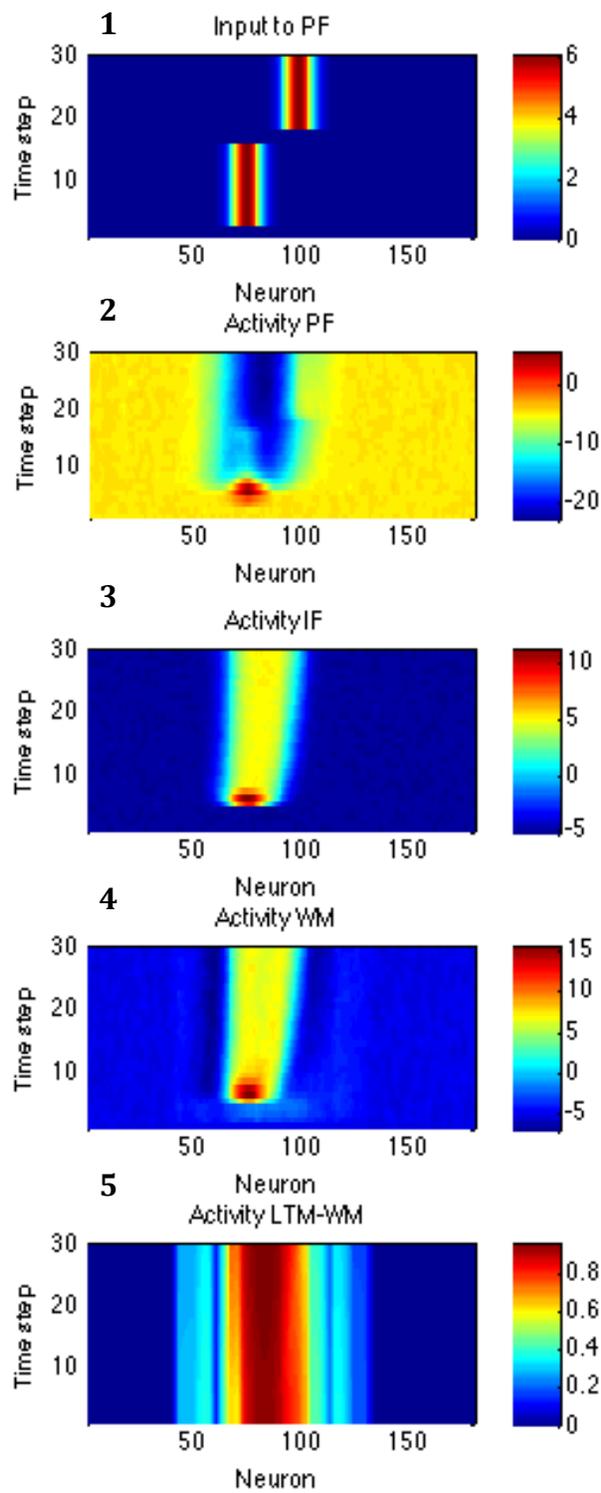


Figure 6: Increased variance without bimodality did not result in discrimination. Plots show activity during Different test trials in the Variance-Control condition. 1) Stimulus 2) PF 3) IF 4) WM 5) LTM-WM

Discussion

The DFT model reproduces the distributional categorization effect. Rather than forming explicit, discrete categories, the drifting of internal representations occurs so that they tend to gravitate toward regions of the perceptual space with more experience (i.e. stronger LTM activity). This has the effect of bringing representations closer to the most typical member of the category, which increases perceived similarity within a category and decreases perceived similarity between categories. Limitations of the model include that here, as in laboratory experiments, the experience is presented all in one continuous block. In the actual environment of developing organisms, while rich distributional experience is available, it is interspersed with many other experiences over a long period. Thus, for this mechanism to operate over developmental time in infants, there must be a way to maintain multiple partially learned distributions simultaneously.

Another limitation is the nature of the model's output. Because the behavioral data from infants consist of looking times that were either the same or significantly different across conditions, the presence or absence of a peak is a sufficient output to evaluate the fit with empirical data. However, it would be useful to predict quantitatively the degree to which within-category stimuli are perceived as more similar and between-category stimuli are perceived as more different. This could be reflected in the model as a function of the size or duration of peaks rather than their mere presence, and could be measured.

The model architecture has been applied in nonlinguistic cognitive tasks such as spatial recall, spatial discrimination and reaching [17, 18, 19]. Because distributional learning can be captured using general cognitive mechanisms, it should not be specific to categorization of speech sounds. We therefore hypothesized that the same mechanism could be used to categorize dynamic actions. Provided that the dynamic visual environment provides similar distributional regularities to the static objects and speech sounds infants encounter, they should be learnable by similar mechanisms. However, although distributional learning of speech and static visual displays has been observed [9, 10, 14], dynamic visual displays have not been studied in this way. To test this hypothesis we designed an experiment in which the stimuli had the same distributional structure, but consisted of animated videos of actions.

Study 2: Infant Learning

Methods

Participants

Thirty-seven 10.5- to 12-month-old infants participated in the study. Participants were recruited through birth announcements in the local newspapers. Infants were randomly assigned to the unimodal or bimodal condition (Unimodal condition: $n = 20$, 6 female, 14 male, age range 332 – 374 days, mean = 347; Bimodal condition: $n = 17$, 9 female, 8 male, age range 322 – 366 days, mean = 347). Infants received an infant t-shirt or bib as a gift for participation.

Stimuli

Each participant viewed a corpus of animations of a bug moving across the screen. There were also three gray boxes on the screen; one at the lower left, one in the center, and one at the upper right. In each familiarization animation, the bug either emerged from the lower-left box and disappeared behind the center box, or emerged from the center box and disappeared behind

the upper right box. The path of the bug was identical except that it occurred with equal frequency on opposite sides of the screen.

The manner of the bug's motion differed along a continuum between actions. At one extreme, the bug was repeatedly compressed longitudinally and did not rotate. At the other extreme, the bug rotated in alternating directions through 90 degrees and was not compressed longitudinally. Fifteen intermediate stimuli were constructed using linear steps of rotation angle and linear steps of compression, such that as one type of motion increased, the other decreased (Fig. 7). Thus, the actions differed from each other by objectively equal steps, but were designed not to differ in salience.

Manner Stimulus 1



Manner Stimulus 5



Manner Stimulus 11



Manner Stimulus 15



Figure 7: Example stimuli. Pictured are still frame sequences of animations 1, 5, 11, and 15, where 1 and 15 are the extremes.

Familiarization animations consisted of one of these 15 manners, and could have either of the two paths. In the test animations, the bug emerged from the lower-left box, passed behind the center box, and finally disappeared behind the upper right box (Fig. 8). The manner during the second half of the path was either the same as the first half or different. The manners used were steps 4 and 12 of the continuum, because these were seen an equal number of times by participants in both conditions. Thus the test trials came in four forms: 4-4 (same), 12-12 (same), 4-12 (different), and 12-4 (different).

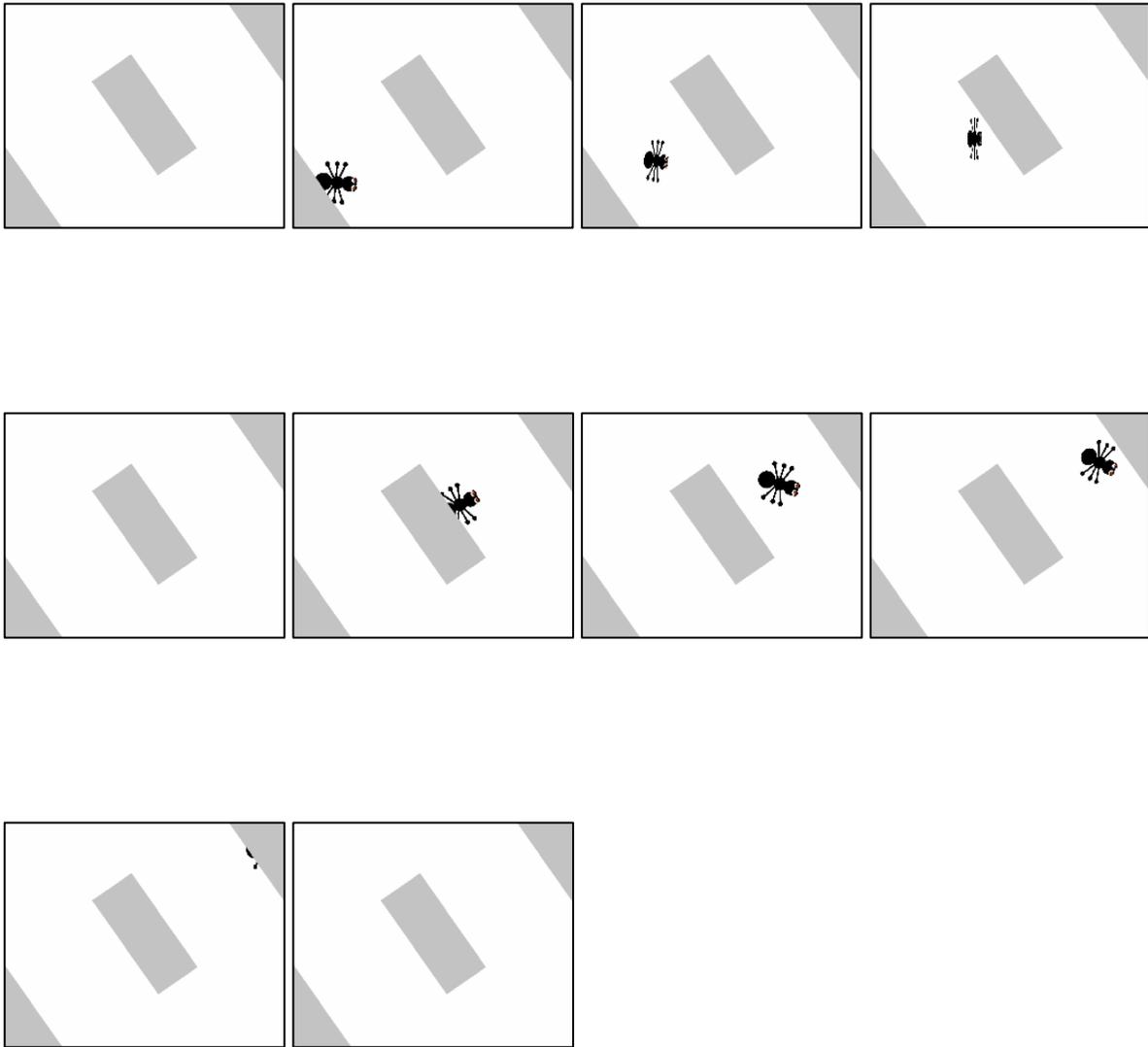


Figure 8: Screen shots from a test trial. The motion before and after the midpoint could either maintain the same manner (Same) or change in manner (Different).

Procedure

All participants were tested by the author. During the familiarization phase, each infant saw a total of 47 animations. Each infant saw each animation, but the number of times each animation was seen differed between conditions. The distributions used contained the same

unimodal/bimodal structure as those in the simulation and Maye et al. (2002), but were modified to include a finer continuum with 15 steps instead of 8 (Fig. 9). They contained 47 total trials instead of 64 to limit the duration of the experiment and prevent fussiness; simulations with this distribution produced all the same results. Each familiarization animation lasted 4 seconds. The order of familiarization trials was random. The side of the screen in which the motion occurred (lower-left or upper-right) was counterbalanced so that each participant saw an equal number of stimuli on the two sides of the screen, and each participant saw manners 4 and 12, which were used in the test trials, same number of times on each side of the screen. Between familiarization and test, infants saw an animation of the bug traversing the same path as in the test trials, but without rotation or compression. This ensured that infants were still attending to the test trials and also pre-exposed infants to the full-screen path that was used during the test trials to reduce its novelty. Then, infants saw two blocks of 4 test trials (two Same, two Different), lasting 8 seconds each. The order of test trials was random within each block.

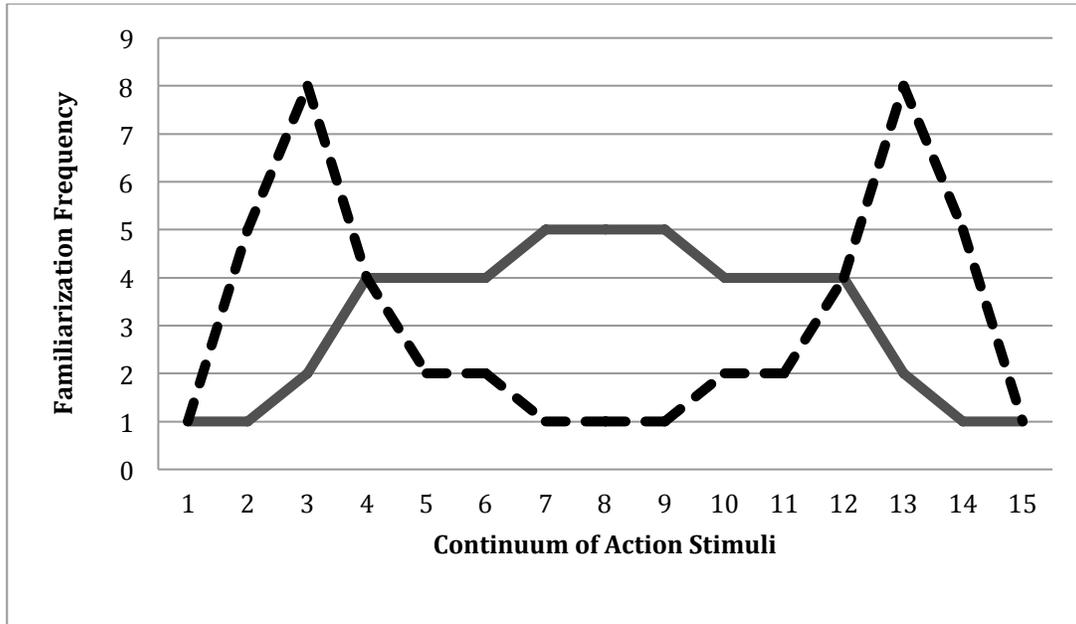


Figure 9: Bimodal vs. Unimodal distributions of stimuli used for familiarization. The distribution for the Bimodal group is shown by the dashed line, and the Unimodal group by the solid line.

Infants watched the familiarization and test phases seated in the caregiver’s lap, facing the screen onto which the stimuli were projected. A camera mounted on the wall faced the infant so that the experimenter in an adjacent room could monitor looking. Caregivers wore a baseball cap with a veil attached so they could see their infants but not the screen.

Each infant saw the same number of trials, with the exception that if the infant did not look at the trial for at least 1 second, the trial was repeated. An attention getting stimulus played between trials, and the next trial started when the infant looked at the screen while the attention getter was playing.

Looks and looks away were coded if they lasted at least 0.5 seconds. Total looking time and number of looks away were recorded for each familiarization trial and test trial. For test

trials, we were especially interested in whether infants tended to look differently at the second half of Same versus Different trials. We hypothesized that infants who discriminated between different motions would perceive the change in motion that occurred in Different trials as a salient event, and would therefore show different amounts of looking to the second half of Same versus Different trials, immediately after the change or non-change. We predicted that only infants in the Bimodal condition would look more at Different trials.

Results

Look-away rate

The experiment was divided into five periods: familiarization trials 1-12, trials 13-24, trials 25-36, trials 37-47, and the test phase. For each participant the look-away rate was computed for each period as the number of looks away divided by total looking time (Fig. 10). An ANOVA with period as within-subjects factor and condition as between-subjects factor revealed a main effect of period, $F(4, 35) = 13.27, p = 0.001$. Post-hoc tests revealed that the look-away rate increased from trials 1-12 to trials 13-24, $t(36) = 4.47, p < .001$, and from trials 13-24 to trials 25-36, $t(36) = 2.54, p = .015$, but not from trials 25-36 to trials 37-47, $t(36) = 1.646, p = .11$. Look-away rate decreased from trials 37-47 to Test, $t(36) = -3.61, p = .001$. There was no significant main effect of condition, $F(1, 35) = 1.32, p = .26$, and no significant period x condition interaction, $F(4, 35) = 0.43, p = .079$.

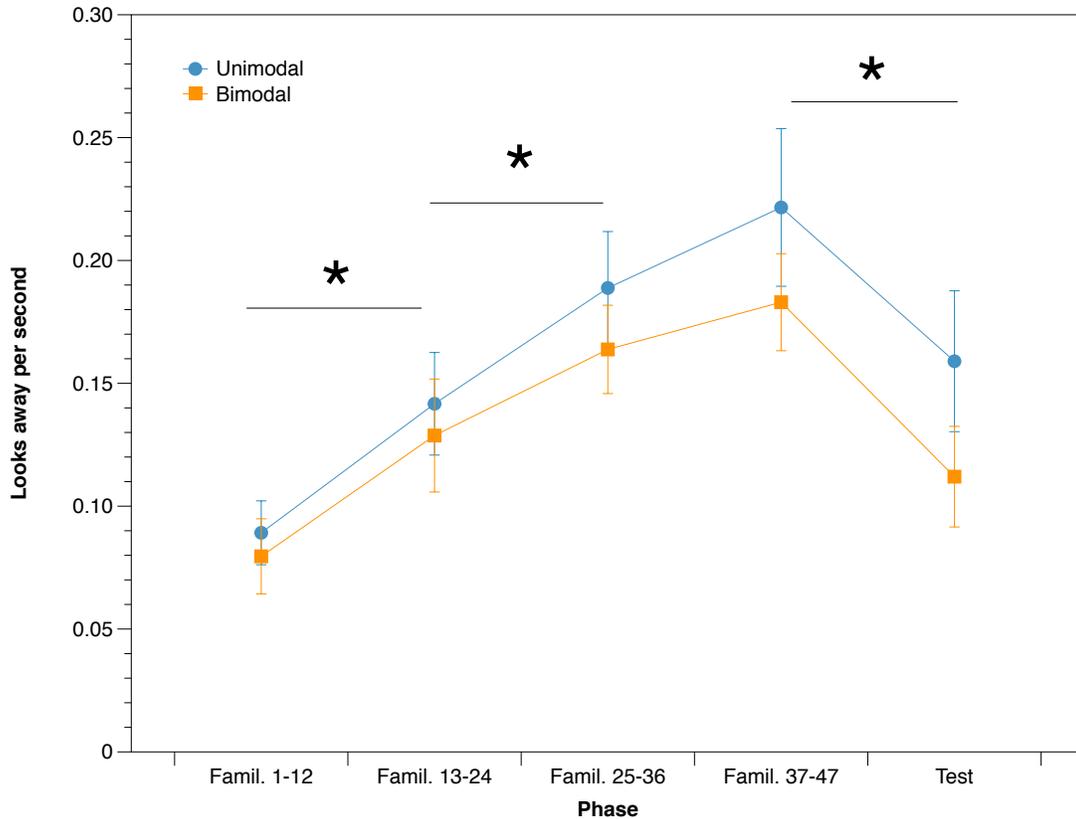


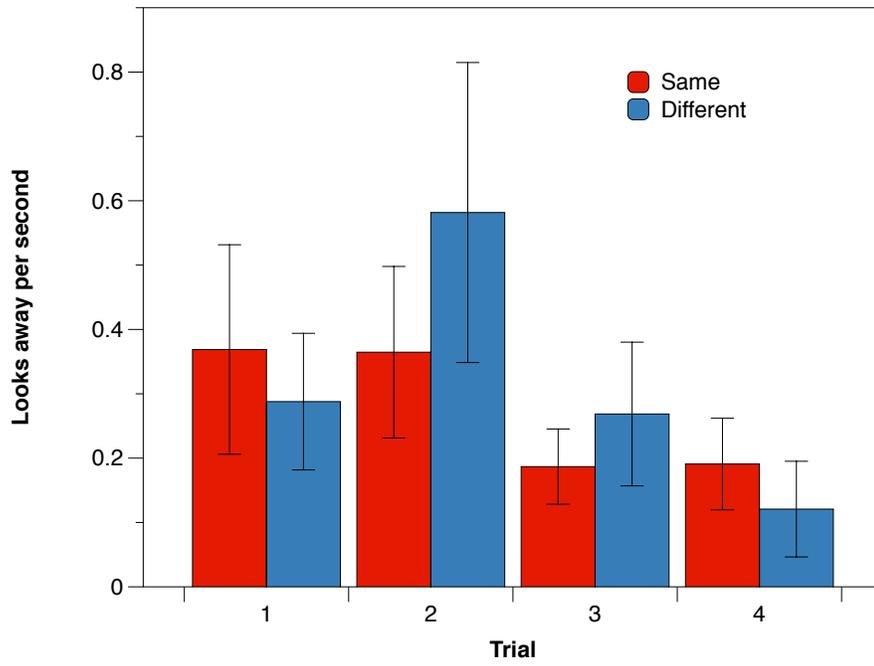
Figure 10: Mean look-away rates (± 1 SE) over the course of familiarization and test phases. Look-away rates increased from trials 1-12 to 13-24 and from 13-24 to 25-36 in the familiarization phase, and decreased from the end of the familiarization phase to the test phase, showing a pattern of habituation and dishabituation.

Test phase

Look-away rates were recorded for the second half of each test trial. For each condition, the look-away rates were compared between Same and Different test trials. In addition, trials were separated by order, which ranged from 1 to 4, i.e. the first to fourth trial of that type seen by that participant (Fig.11). Because the look-away rate for a single trial was non-normally

distributed, means were compared using Wilcoxon signed-rank tests. A pair of Same and Different points were considered matched if they came from the same participant and had the same order. Look-away rate did not significantly differ between Same or Different trials in the Bimodal condition ($p = .853$) or in the Unimodal condition ($p = .438$). No significant differences were found by comparing only trials of a specific order ($ps > .2$).

A Unimodal condition



B Bimodal condition

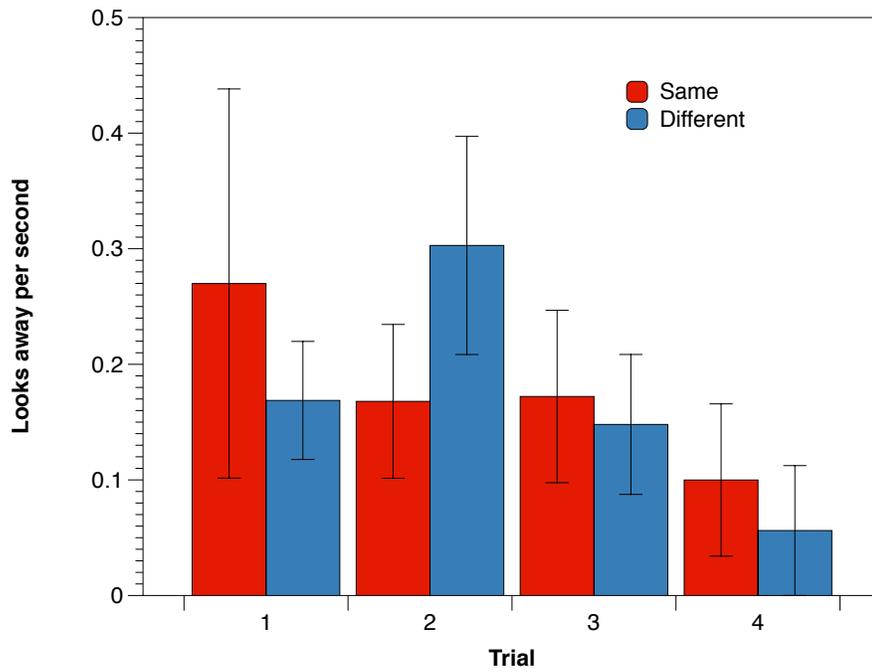


Figure 11: Mean look-away rate (± 1 SE) for Same vs. Different test trials by condition.

A: Unimodal condition. B: Bimodal condition

Discussion

Over the course of the familiarization phase, look-away rate increased in both conditions due to habituation, then decreased again during the test phase as infants dishabituated to the changes in the stimulus. There was a tendency for infants in the Bimodal condition to look away less frequently than infants in the Unimodal condition, which increased over the course of the experiment, though the effect was not statistically significant. It would be interesting to increase the sample size to determine whether conditions would differ with higher statistical power. If so, this would indicate that infants attend to the learnable category structure, as their attentiveness is not different at the start of familiarization but diverges as they gain more familiarity with the distributional properties of the stimulus.

Variability in rate of looking away during test trials was too high to determine whether infants in the different conditions perceived the animations differently. We did not observe evidence that infants in any condition differentiated Same test trials from Different test trials.

A potential future study would test older infants to investigate whether distributional information facilitates verb learning. I would familiarize infants with a unimodal or bimodal distribution of actions, then test whether distributions affect their ability to subsequently learn verbs referring to these actions.

Another potential direction is to test adults' categorical learning of the same stimuli. This would allow more detailed response measures than looking time, and may allow characterization of perceptual differences more subtle than same/different judgments. I will expose adults to the same stimuli used here, then test their perception of the similarity between different pairs of

animations from the set. There is evidence that adults can learn categorical perception by distributional learning [14]. Additionally, the DFT model predicts that the effect of distributional learning should increase with a delay between presentation of the two stimuli, as internal representations drift. Therefore, I will test adults' similarity judgments with and without this delay. If significant learning effects are observed in this study, then it will be useful to find a more sensitive measure for learning in infants. Maye et al. [9] used looking time to infer distributional learning in 8-month-old infants. Their test trials differed from ours in that they contrasted Alternating test trials, in which two different stimuli were alternately played, with Non-Alternating trials, in which the same stimulus was repeatedly played. In contrast, our Different and Same test trials each consisted of a single stimulus pair. This was done to limit the duration of the test phase, but could obscure effects if differences in looking emerge on a slow timescale, such that total looking over a series of Alternating and Non-Alternating stimuli may differ while look-away rates immediately after changes to the stimulus may not.

Alternatively, it may be necessary to use additional, more interactive or fine-grained methods of measuring infant looking behaviors [25]. Gaze tracking would allow measurement of looks to and looks away from more precise regions of interest. 6-to-8-month infants can also rapidly learn to look at on-screen "buttons" to produce a contingent response [26]. This type of paradigm could be used to directly test infants' ability to discriminate different motions using a gaze-contingent reward for correct responses. Another related approach is to have different motions predict the appearance of a rewarding visual stimulus and use predictive looking, or looks to the location of the reward before it appears, as a learning measure [27].

Taken together, this program of research will have significant implications for understanding how infants learn to interpret the world. A statistical learning approach to the

study of language acquisition has revolutionized our understanding of the mechanisms underlying language development. [3, 4, 9]. A similar approach to social development can have a similar effect on how infants learn to interpret others' social behavior. Additionally, this knowledge can inform language and social interventions in young children, as well as artificial intelligences that learn to identify and categorize relevant items in a complex input. While clustering algorithms exist to solve categorization problems, it is not known how corresponding processes are implemented in the brain, or how they are linked to other cognitive computations.

In addition, further research into this and related questions will help elucidate how learning mechanisms and structural regularities in social behavior coevolve. This idea has been well elaborated in the case of the structure of language coevolving with the capacities of the human brain [28], but, just as the statistical learning processes that support language acquisition, it is likely to be far more general. For example, the regularities present in people's non-linguistic gestures and other social actions may be influenced by infants' ability to perceive, attend to, and learn that structure. Comparative studies could investigate how distributional or sequential structure in social actions in different species of birds, rodents or nonhuman primates relates to the learning abilities of those species.

Appendix

Model Equations

The neural field dynamics used here were originally introduced by Amari [16]. For a developmental application of a similar model, see [29].

Activity in PF changes according to Equation 1:

$$\tau \dot{u}(x, t) = -u(x, t) + h_u + S(x, t) + q\xi(x, t) + \int w_{uu}(x - x')f(u(x'))dx' - \int w_{uv}(x - x')f(v(x'))dx' \quad (1)$$

$u(x, t)$ represents the activity of neuron x at time t . τ defines the time course of the field's dynamics. The $-u(x, t)$ and h_u terms together cause the field's activity to decay to resting level h_u . $S(x, t)$ represents the input to neuron x and time t . This is determined as described below from the activity of the other fields and, for PF, the stimulus. $q\xi(x, t)$ is Gaussian distributed noise, $\int w_{uu}(x - x')f(u(x'))dx'$ represents the cooperative interactions among locations in the field, and $-\int w_{uv}(x - x')f(v(x'))dx'$ represents inhibition from IF. The *interaction kernel*, $w(x - x')$, specifies the magnitude of interaction between two neurons, x and x' , and takes the form of a Gaussian with mean 0 so that nearby sites interact cooperatively (Equation 2). f is a sigmoid soft threshold function applied so that only neurons with significant activity influence other neurons (Equation 3). Integration over all values of x' provides for input from all active sites. Finally, $S(x, t)$ is the input to neuron x at time t from the stimulus.

The interaction kernel w consists of a scaled Gaussian:

$$w(x - x') = c \exp\left[-\left(\frac{(x-x')^2}{2\sigma^2}\right)\right] \quad (2)$$

The sigmoid function is given by:

$$f(u) = \frac{1}{1 + \exp[-\beta u]} \quad (3)$$

The input is given by

$$S(x, t) = c \exp \left[- \left(\frac{(x - x_{center})^2}{2\sigma^2} \right) \right] \quad (4)$$

for timesteps in which a stimulus is present, where x_{center} varied from 58 to 114 in steps of 8, and $S(x, t) = 0$ when no stimulus is present. Each trial lasted 225 timesteps and the stimulus was present for steps 20 – 220.

Activity in IF is computed according to Equation 5:

$$\begin{aligned} \tau \dot{v}(x, t) = \\ -v(x, t) + h_v + q\xi(x, t) + \int w_{vu}(x - x')f(u(x'))dx' - \int w_{vw}(x - x')f(w(x'))dx' \end{aligned} \quad (5)$$

IF receives input from both PF and WM, but not from other sites in IF.

Activity in WM is computed according to Equation 6:

$$\begin{aligned} \tau \dot{w}(x, t) = \\ -w(x, t) + h + q\xi(x, t) + \int w_{wu}(x - x')f(u(x'))dx' - \int w_{wv}(x - x')f(v(x'))dx' + \\ \int w_{wu}(x - x')f(u(x'))dx' + \int w_{w_ltm}(x - x')f(u_{ltm}(x'))dx' \end{aligned} \quad (6)$$

WM receives excitatory input from itself, from PF, and from LTM-WM, and inhibitory input from IF.

Activity in LTM-WM is computed according to Equation 7:

$$\dot{u}_{ltm}(x, t) = \frac{-u_{ltm}(x,t) + f(w(x,t))}{\tau_{build}} I_{w(x,t) > 0} + \frac{-u_{ltm}(x,t)}{\tau_{decay}} (1 - I_{w(x,t) > 0}) \quad (7)$$

where $I_{w(x,t) > 0} = 1$ if $w(x, t) > 0$, and 0 otherwise. Thus whenever a neuron in WM has positive activity, its output is added to LTM-WM. LTM-WM has two time constants τ_{build} and τ_{decay} , used for neurons that receive input on that timestep or receive no input, respectively.

Model Parameters

Time constants:

$$\tau_u = 20, \tau_v = 20, \tau_w = 20, \tau_{build} = 1000, \tau_{decay} = 5000$$

Resting levels:

$$h_u = h_v = h_w = -5$$

Slope of sigmoid function:

$$\beta = 5$$

Interaction kernels:

$$c_{uu} = 25, \sigma_{uu} = 5$$

$$c_{uv} = 20, \sigma_{uv} = 10$$

$$c_{vu} = 10, \sigma_{vu} = 5$$

$$c_{wu} = 20, \sigma_{wu} = 5$$

$$c_{wv} = 20, \sigma_{wv} = 10$$

$$c_{ww} = 25, \sigma_{ww} = 5$$

$$c_{w_ltm} = 4, \sigma_{w_ltm} = 2$$

References

- [1] Christiansen, M., Onnis L., & Hockema, S. (2009). The secret is in the sound: from unsegmented speech to lexical categories. *Developmental Science*, *12*, 388-395.
- [2] Soderstrom, M., Conwell, E., Feldman, N., & Morgan, J. (2009). The learner as statistician: three principles of computational success in language acquisition. *Developmental Science*, *12*, 409-411.
- [3] Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926-1928.
- [4] Misyak, J. Goldstein, M., & Christiansen, M. (2012). Statistical-sequential learning in development. In Rebuschat, P & Williams, J.N. (Eds.) *Statistical Learning and Language Acquisition*, pp. 13-54. Boston: Walter de Gruyter.
- [5] Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*, B35-B42.
- [6] Pruden, S., Göksun, T., Roseberry, S., Hirsh-Pasek, K., & Golinkoff, R. (2012). Find your manners: how do infants detect the invariant manner of motion in dynamic events? *Child Development*, *83*, 977-991.
- [7] Loucks, J. & Baldwin, D. (2006). In Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). *Action Meets Word* 228-258. Oxford University Press.
- [8] Lewkowicz, D. & Ghazanfar, A. (2006). The decline of cross-species intersensory perception in human infants. *PNAS*, *103*, 6771-6774.

- [9] Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, B101-B111.
- [10] Maye, J., Weiss, D., & Aslin, R. (2008). Statistical phonetic learning in infants: facilitation and feature generalization. *Developmental Science*, 11, 122-34.
- [11] Mareschal, D. & Quinn, P. (2001). Categorization in infancy. *Trends in Cognitive Sciences*, 5, 443-450.
- [12] Mareschal, D. & French, R. (2000). Mechanisms of categorization in infancy. *Infancy*, 1, 59-76.
- [13] Younger, B. (1985). The segregation of items into categories by ten-month-old infants, *Child Development*, 56, 1574-1583.
- [14] Gureckis, T. & Goldstone, R. (2008). The effect of the internal structure of categories on perception. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society*, Austin, TX. Cognitive Science Society. 1876-1881.
- [15] Spencer, J.P., Perone, S., & Johnson, J.S. (in press). The dynamic field theory and embodied cognitive dynamics. In J.P. Spencer, M.S. Thomas, & J.L. McClelland (Eds.) *Toward a Unified Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered*. New York: Oxford University Press.
- [16] Amari, S. & Arbib, M. A. (1977). Competition and cooperation in neural nets. In Metzler, J., (Ed.). *Systems Neuroscience*. Academic Press; New York.

- [17] Thelen, E., Schöner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: a field theory of infant perservative reaching. *Behavioral and Brain Sciences*, 24, 1-86.
- [18] Schöner, G. & Thelen, E. (2006). Using dynamic field theory to rethink infant habituation. *Psychological Review*, 2, 73-99.
- [19] Simmering, V.R. & Spencer, J.P. (2008). Generality with specificity: The dynamic field theory generalizes across tasks and time scales. *Developmental Science*, 11, 541-555.
- [20] Johnson, J.S., Spencer, J.P., & Schöner, G. (2008). Moving to higher ground: The dynamic field theory and the dynamics of visual cognition. In F. Garzón, A. Laakso, & T. Gomila (Eds.) Dynamics and Psychology [special issue]. *New Ideas in Psychology*, 26, 227-251.
- [21] Sokolov, E. (2012). *Perception and the Conditioned Reflex*. New York: Pergamon
- [22] Gerken, L., Balcomb, F. K., Minton, J.L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, 14, 972-979.
- [23] Kidd, C., Piantadosi, S.T., Aslin, R.N. (2012). The Goldilocks effect: human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS One*, 7(5): e36399.
- [24] McMurray, B., Aslin, R. N., & Toscano, J.C. (2009). Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science* 12:3, 369-378

- [25] Aslin, R. N. (2007). What's in a look? *Developmental Science* 10, 1, 48–53.
- [26] Wang, Q., Bolhuis, J., Rothkopf, C.A., Kolling, T., Knopf, M., Triesch, J. (2012). Infants in control: rapid anticipation of action outcomes in a gaze-contingent paradigm. *PLoS ONE*, 7(2): e30884.
- [27] Yurovsky, D., Boyer, T.W., Smith, L.B., & Yu, C. (2013). Probabilistic cue combination: less is more. *Developmental Science* 16:2, 149-158
- [28] Christiansen, M.H. & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, 31, 489–558.
- [29] Simmering, V. R., Spencer, J. P, & Schutte, A. R. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Res.*, 1202, 68–86.