Berthouze, L., Kaplan, F., Kozima, H., Yano, H., Konczak, J., Metta, G., Nadel, J., Sandini, G., Stojanov, G. and Balkenius, C. (Eds.)
Proceedings of the Fifth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems
Lund University Cognitive Studies, 123. ISBN 91-974741-4-2

# Emerging Linguistic Functions in Early Infancy

**Francisco Lacerda*, Eeva Klintfors, Lisa Gustavsson**
**Ellen Marklund and Ulla Sundberg**
Department of Linguistics
Stockholm University
SE-106 91 Stockholm
Sweden
* frasse@ling.su.se

## Abstract

This paper presents results from experimental studies on early language acquisition in infants and attempts to interpret the experimental results within the framework of the Ecological Theory of Language Acquisition (ETLA) recently proposed by (Lacerda et al., 2004a). From this perspective, the infant's first steps in the acquisition of the ambient language are seen as a consequence of the infant's general capacity to represent sensory input and the infant's interaction with other actors in its immediate ecological environment. On the basis of available experimental evidence, it will be argued that ETLA offers a productive alternative to traditional descriptive views of the language acquisition process by presenting an operative model of how early linguistic function may emerge through interaction.

## 1. Introduction

Previous studies of the young infant's ability to learn names of objects presented under controlled naturalistic settings have demonstrated that by about 7 to 8 months of age, infants are capable of interpreting arbitrary words as names of visual objects, provided the words and the objects co-occur consistently. For instance, a study by Gogate and Bahrick (Gogate & Bahrick, 1998) indicates that 7 month-old infants are able to explore audio-visual co-occurrences to establish arbitrary word-like associations between isolated speech sounds and objects. In addition, a more general assessment of the impact of audio-visual synchrony (Prince et al., 2004) strongly suggests that synchronic events may expose linguistically relevant audio-visual relationships. But while the young infant's ability to establish sound-object links offers god support to the notion that association processes are likely to underlie early language acquisition, accounting for the language acquisition process in terms of relatively simple associative processes involving isolated words is problematic because it may lack general ecological relevance. Indeed, as often pointed out by scientists criticizing the emergentist views of the language acquisition process (Lidz, Gleitman, & Gleitman, 2003), words representing the names of objects available to the young infant tend to be embedded in utterances rather than uttered in isolation, an aspect that necessarily reduces the ecological relevance of experimental studies reporting referential learning from words presented in isolation. Thus, to further investigate the extent to which general association processes might underlie early language acquisition in ecologically relevant adult-infant interaction settings, a series of experiments were set up in which the target words were integrated in natural sentences (as those typically heard by infants) and arbitrarily combined with visual objects simultaneously accessible to the infants.

The present paper will argue that early language acquisition can indeed be seen as the result of an interactive process between the infant and its environment, through which the infant picks up linguistic regularities afforded in the ambient language. In the following we try to provide an empirical basis for our emergentist views of the early language acquisition process by reviewing some of our experimental studies addressing different aspects of early language acquisition in infants and examining the characteristics of the infant-directed. We will first review an experiment designed to test how different linguistic factors may influence the infant's ability to derive word-object relationships from exposure to naturalistic audio-visual contingencies. Thereafter we will examine the characteristics of infant-directed speech from the perspective of the ETLA. Finally, we will address the issue of necessity of general-purpose versus language-specific processes underlying the infant's ability to link visual and auditory information and form productive linguistic representations.

## 2. Emerging word-object associations

To investigate the generality of the word-object association process, a series of experiments were carried out to investigate the infant's ability to derive the names of objects from experience with audio-visual stimuli, where natural sentences conveying implicit referential information are presented simultaneously with the visual images of the objects they refer to. One setup of these studies was already described in (Gustavsson et al., 2004) and will be briefly reviewed here.

This study used a Visual Preference procedure similar to the procedures used by Fernald and her colleagues (Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999). In general terms, the procedure can be described as inducing the infant's response from its looking time towards alternative pictures displayed simultaneously and where one of the pictures is associated with the expected response.

## 2.1 Speech materials

The speech materials were Swedish sentences recorded by a female native speaker of Swedish. The uttrances introduced non-words as names of the objects being displayed on the screen. Nine films were created to include all the possible combinations of position of the target word (initial, medial or final position in the utterance) and the utterances focal accent (falling on the utterances initial, medial or final words). The syntactic structure of the utterances was different from film to film but within each film the position of the target word and the part of the utterance receiving focal accent was kept constant. Furthermore, although the utterances within each film were structurally equal, the non-target words were different from utterance to utterance in an attempt to mimic the variation typically observed in natural utterances. Examples of the utterances presented in two of the nine films are shown in table 1, where the focal accent is indicated by boldface and the position of the target word by XXX. For the placement of the target word and focal accent, the utterances were divided in three regions – initial, medial and final. The initial and final positions were defined by the first and the last word in the utterance. The medial position was defined as the remaining part of the utterance.

| Film 1<br>target word: final<br>focal accent: final | Film 2<br>target word: final<br>focal accent: medial |
|---|---|
| Titta här är söta **XXX** | Det **är** den söta XXX |
| Se på den lilla **XXX** | Se på den **lilla** rara XXX |
| Titta på fina **XXX** | Titta på **fina** glada XXX |
| Kolla in den glada **XXX** | Kolla **glada** XXX |

Table 1. Example of the Swedish utterances presenting the target words. The target word is represented by XXX, standing for the non-words "Kucka" and "Dappa". Focal accent is represented by boldface.

Each of the nine films was organized in three phases – baseline, exposure and test.

In the baseline phase, still images of two puppets were displayed side by side in a split-screen. The duration of the baseline phase was 30 s. During the baseline phase an especially composed short instrumental lullaby (Anna Ericsson, 2004) was played, starting approximately 2 s after the onset of the visual display and finishing about 2 s before the end of the baseline phase. The infant's looking towards each of the puppets during this phase was used a measure of the subject's preferential bias towards the puppets.

During the exposure phase, two short 20 s video sequences were played to show each of the puppets per se, introduced by the sentences referring to the particular puppet being displayed (see table 1). The sentences were evenly distributed throughout the duration of each video sequence. The first sentence started about 1 second after the onset of the visual display and the last sentence finished about 1 s before switching to the next video sequence. These video sequences were presented after each other, switching from one puppet to the other. The total duration of the exposure phase was 120 s, during which each of the individual video sequences was presented 3 times. The infants' looking time towards the each of the puppets was taken as a measure of attention during the exposure phase.

In the test phase the two puppets were again displayed in a split-screen similar to that of the base-line but now the audio track played questions like "Where is XXX?" or "Can you see XXX?", where XXX was the name of one of the puppets, implicitly introduced in the descriptions presented during the exposure phase. The test phase was 30 s long, just as the baseline phase.

## 2.2 Subjects

A total of 49 infants participated in the study. Some of the infants participated in more than one session, adding up to 78 sessions. The results presented here come from a total of 75 sessions, distributed as indicated in the table below. The ages of the subjects at the time of their participation in the sessions ranged from 201 to 278 days (mean age was 239 days, s.d.=15 days). The age distribution for this sample was nearly Gaussian (skewness=0.180; kurtosis=0.503).

| | | Focal accent | | | |
|---|---|---|---|---|---|
| | | initial | medial | final | Total |
| Target word position | initial | 10 | 7 | 8 | 25 |
| | medial | 10 | 7 | 10 | 27 |
| | final | 8 | 8 | 7 | 23 |
| | Total | 28 | 22 | 25 | 75 |

Table 2. Number of data points used for specific combinations of target word and focal accent positions.

The selection criterion for the sessions above required that the looking bias towards one of the puppets did not exceed 35% of the baseline time.

## 2.3 Procedure

The subjects were video recorded during the experiments, using a camera placed just above the display they were looking at. To register the actual images that the infant was looking at and give the possibility of re-analysis, the film being displayed was mixed onto the upper left corner of the image of the subject's face. This overlapping image used about 1/16 of the screen area and did not interfere with the image of the face of the infant. A time stamp placed at 40 ms intervals was also recorded on the lower right corner of the screen. This time stamp was subsequently used to compute a session-relative time, allowing the line up the start of the video films from different subjects.

In this experiment, the looking times towards each of the puppets were measured manually, frame by frame, a very time consuming procedure. The separation between the target images was about $30^o$, which was enough to allow clear decisions on which side of the screen the subject was momentarily looking at. Three levels of looking were coded – left, right and off.

On the basis of these codes, a "pre-to-post exposure gain" variable was defined as the net increment in looking time towards the puppet used as target:

$$Gain = Tgt - TgtB$$

where Tgt is the total looking time towards target puppet and TgtB is the total looking time during baseline towards the puppet that would become the target in the test phase.

## 3. Results

The results from the first sessions in which the 36 selected subjects participated are shown in figure 1, grouped according to the placement of the focal accent and the position of the target word in the utterances.

Given the reduced number of subjects each condition and the typical variance observed in this type of experiments, it is perhaps not surprising that no significant main effects for the target word or placement of the focal

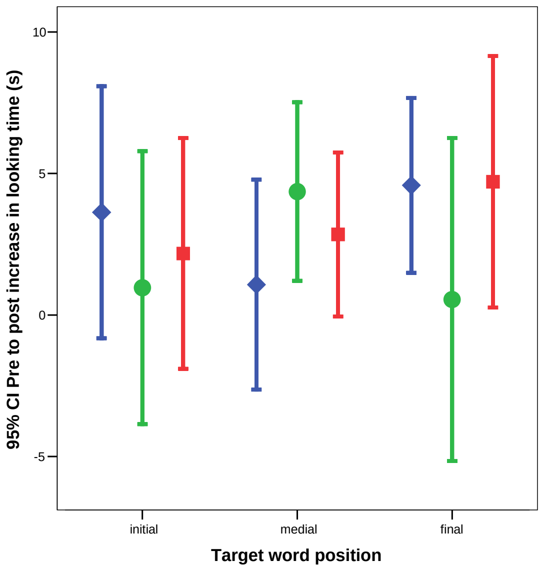


Figure 1. Average gains (in seconds) and respective 95% confidence intervals for the infants' responses as a function of the placement of the target word and focal accents (diamonds represent focal accents on the initial word, circles on medial words and squares focal

accent could be observed. As stated above, this analysis was carried out on a selection of all the sessions in which the infant's initial bias towards any of the puppets was less than 35% of the total baseline time. Further analyses using all the available data from the 78 sessions did not change appreciably the pattern displayed on figure 1. The main difference was a broadening of the confidence interval for medial target word position with medial focus, due to an extreme negative gain outlier resulting from a strong bias towards the puppet that would function as target. A non-parametric display of the same data is shown in figure 2. The dependence of the median values on the target word and focal accent position is in good agreement with the pattern displayed in figure 1. There were no significant main effects or interactions for word position and placement of the focal accent. However a tendency for longer looking times was observed for the target word in focal position ($F(1,73)=2.957$, $p<0.090$). If the case of the target word in final position, with a focal accent in the initial position of the sentence is excluded,

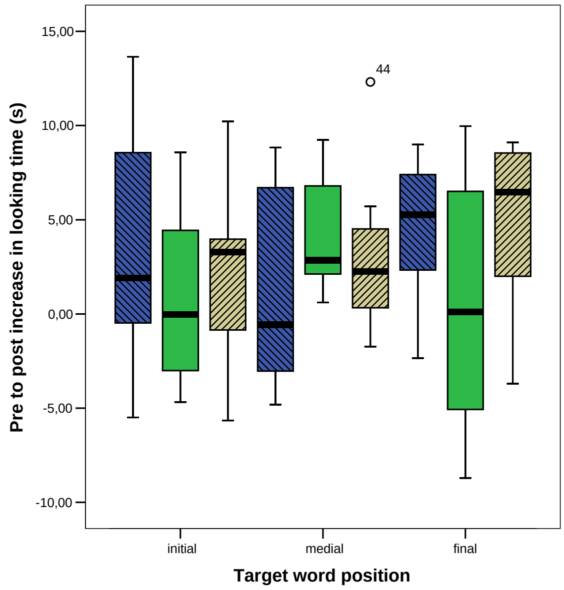


Figure 2. Box plots of the data displayed in figure 1. The order of the placements of the focal accents is also the same as in figure 1.

then a significant effect of the placement of the target word in focus is obtained ($F(1,65)=4.075$, $p<0.048$). Furthermore, there was a significant difference between the mean looking times for the group of sentences with the target word in focal position *plus* the sentences with the target word in final position and focal accent in initial position ($F(1,73)=5.579$, $p<0.021$).

## 4. Discussion

While there were no overall significant differences when considering all the data at once, the response pattern displayed in figures 1 and 2 strongly suggested that target words in focal position might have been easier to associate with the corresponding puppets than when focal accent did not fall on the target word. This means that 8-month-old infants seem to be able to pick up relevant linguistic information by listening to the word that is placed in focal position. An unexpected effect was however observed when the target word was in non-focal final position but the utterance had initial focal accent. It appears that the initial focal accent may have primed the infants to attend to the utterance, prompting the subjects to retrieve the less prominent target word delivered in sentence final position.

In summary, the results of this experiment seem to indicate a general ability to link recurrent target words with visual objects that are simultaneously available to 8-month-old infants, providing the ground for the linguistically relevant referential function. The fact that the strength of the responses varied significantly for different combinations of focal accent target word placement further suggests that the infants' ability to pick up the linguistic referential function was modulated by prosodic patterns and primarily contingent on the coherence in the placement of the focal accent and the target word. An implication of this is that by 8 months of age, deriving a linguistic referential function on the basis of exposure to running speech may not be simply a matter of co-occurrence of recurrent sound strings (representing the target word) and visual objects (the puppets to which the target words refer), rather a process in which the acoustic salience of the sound strings plays a decisive role. This notion is reinforced by the fact that focal accent on the initial part of the utterance seems to have enhanced the response to the target word that occurred in utterance final position.

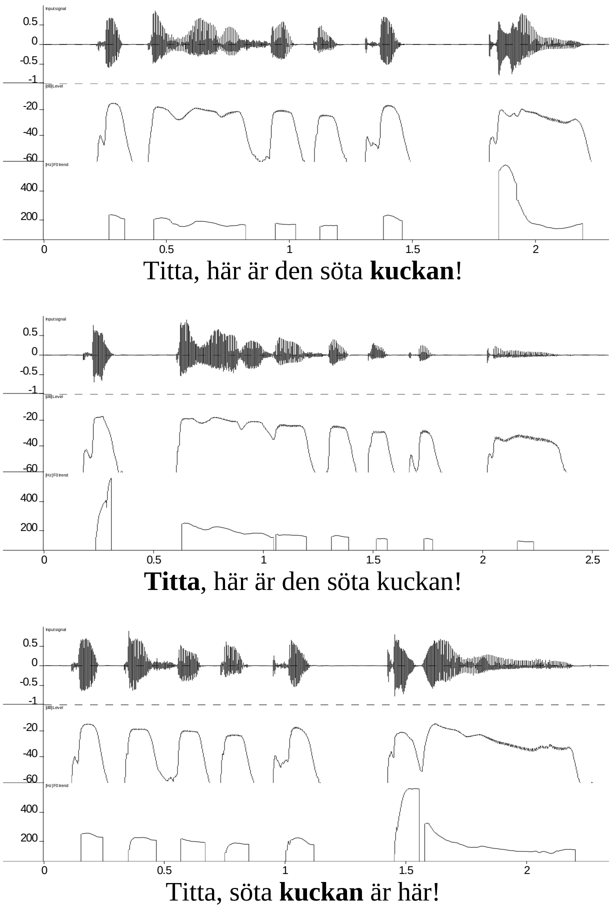


Figure 3. Waveforms (-1 to +1), intensity curves (-60 dB to 0 dB) and $f_0$ contours (100 Hz to 600 Hz) for the first sentences in the conditions. Time scale in seconds.
Top – Focal accent and target word in final position
Mid – Focal accent initial, target word final Bottom – Focal accent and target word medial. The target word is uttered during the interval 0.9-1.6 s, approximately. The high and flat $f_0$ contour at about 1.5 s in this sentence is not due to saturation.

Three examples of sentences introduction the target word “kucka” are shown in figure 3 to illustrate the dramatic $f_0$ excursions associated with the focal accent. Such pitch variations are typical of infant-directed speech. They introduce over one octave increase in pitch and are likely to be salient enough to capture the infant’s attention towards the sound string being uttered. Our previous studies addressing younger (under 6 months of age) infants’ to discriminate between short utterances exposing the interaction between the target words and focal accent placement suggested however that those younger infants might not be able to pay attention to the target word when strong focal information was added (Lacerda & Sundberg, 1996). To be sure Lacerda and Sundberg’s (1996) study concerned discrimination and the speech materials were much simpler than those used in the experiment above, but the results may probably be taken as an indication that the younger infants might not have succeeded in the current task.

# 5. Linking audio-visual information

The emergence of the linguistic referential function suggested by the study reported in the previous section may be seen as a consequence of a general multi-sensory representation process through which synchronic multi-sensory information is spontaneously associated, thereby exposing implicit cross-modal regularities (Lacerda et al., 2004a; Lacerda et al., 2004b; Lacerda, 2003). Because the efficient use of spoken language is based on the ability to relate sound symbols (however variable) to objects perceived (primarily but not exclusively) by other senses, a systematic (or at least predictable) link between the sound code and the objects it refers to must exist at some level of representation (Minsky, 1985). Note that in line with ETLA (Lacerda et al., 2004a), such a sound code is a generic reference to the concrete auditory impression of a word or a lexical phrase as a whole, not to the word’s representation in terms of linguistic concepts like phonemes or syllables nor to the sequence of words that may build up the lexical phrase. In this perspective, words, syllables and phonemes are an emergent consequence of the combinatorial pressure imposed by increasing representation needs (Nowak, Plotkin, & Jansen, 2000; Lacerda, 2003).

To address the issue of the generality of cross-modal links in infancy, we carried out a study to investigate the infant’s ability to use temporal synchrony to relate ecologically relevant auditory and visual speech information, the infant’s ability to relate ecologically relevant synchronic non-speech audio and visual information and the infant’s ability to relate synchronic non-speech audio with speech (articulatory) visual information.

The background for the present experiments is an early Kuhl and Meltzoff’s study (Kuhl & Meltzoff, 1982) showing that 18 to 20 weeks-old infants can pick up the correlation between acoustic and articulatory characteristics of speech sounds. In their study the infants were exposed to a split-screen displaying two faces, one articulating [a] and the other articulating /i/, while an audio signal consisting of either one of those vowels was played. Their results indicated significantly longer looking times towards the face whose articulation was consistent with the audio signal.

Also (Bahrick, 2004) carried out a study in which 5 months-old infants were tested on their ability to discriminate between different phenomena involving changes in rhythm or tempo. The tests were organized in three situations: (1) a multimodal situation, where a plastic hammer was seen while the sound of the hammer hitting a surface was heard, (2) a unimodal situation where only the sound of the hammer was heard, and (3) a unimodal situation where the hammer was only seen but not heard. The rhythm of the events was subsequently manipulated in each of these three situations in order to create novel situations that the infants might discriminate. The results indicated that only the 5 month-old infants who received the bimodal redundant stimulation could detect the rhythm changes. According to other studies by Bahrick, infants tend to be less dependent on redundant information the older they get.

Our study attempted to expand the findings of Kuhl and Meltzoff (Kuhl et al., 1982) and Bahrick's by investigating the ability of 6 to 8 months-old Swedish infants to perceive synchronous visual and auditory input, for both speech and non-speech events. We also introduced a methodological improvement by using a high resolution eye-tracking system, with a maximum resolution of about 0.5°, which allowed the presentation of four images on a single screen during the test phase instead of the two alternatives used by Kuhl and Meltzoff, thereby reducing to 25% the spontaneous chance level of looking at one of the images. Just as in Kuhl and Meltzoff's case, we hypothesized that the infants would look significantly longer towards the images displaying motor activities coherent with the heard speech or non-speech signals.

# 6. Method

After a short calibration of the eye-tracking system, the infants were exposed to a short video film while their eye-movements were registered throughout the session. For this paper, only the infants' average looking times towards the different quadrants of the split-screen will be considered for statistical analysis. However, the eye-tracking data was collected with high enough temporal and spatial resolution to allow a detailed study of the infants' visual strategies but those results will be reported in a future paper.

## *6.1 Subjects*

Of the forty infants who participated in this study four had to be excluded due to calibration errors. The resulting in 36 subjects (13 boys and 13 girls) aged 25-33 weeks (mean age 28.5 weeks). The subjects were randomly selected from the National Swedish address database (SPAR) targeting 6 to 8 months-old infants whose parents lived in the Stockholm metropolitan area.

## *6.2 Stimuli*

The infants were exposed to a film showing a female actress against a blue background. The film consisted of four sequences: (1) a baseline for speech articulations, (2) a test phase for audio-visual coherence in speech stimuli, (3) a baseline for non-speech gestures and (4) a test phase for audio-visual coherence in non-speech stimuli.

In the speech part or the experiment the baseline consisted of four identical still images on a split-screen showing the actress's face. The baseline of the non-speech part of the experiment was an animated video sequence showing four different tempos of hand clapping, one in each quadrant. This baseline sequence was identical to the one to be used in the test phase, but with a silent sound track.

In the test phase for the speech stimuli (figure 4), the actress was again shown on a four quadrant split-screen articulating the vowels [a] and [y] and the syllables [ba] and [by]. For periods of 20 seconds, the speech signal was synchronized with the film shown in one of the quadrants.

In the first 20 seconds the speech signal consisted of repetitions of the syllable [by] and the organization of the four video tracks was (target position in boldface)

| | |
|---|---|
| ba | **by** |
| y | a |

Figure 4. Example of the speech sound part of the experiment. The actress is articulating [ba] (UL), [by] (UR), [y] (LL) and [a] (LR). The audio played was the syllable [by], i.e. the target image was UR.

Directly after the vowel [a] was presented in the next 20 seconds and the position of the visual targets was

| | |
|---|---|
| by | ba |
| **a** | y |

After this the [by] syllable was repeated as target but the visual target was placed in another quadrant

| | |
|---|---|
| y | a |
| ba | **by** |

Finally the [a] was presented again and the visual target once more relocated

| | |
|---|---|
| **a** | y |
| by | ba |

The utterances were produced with rise-fall $f_0$ contours and the target images were placed in different quadrants for each of the four 20 seconds sequences, as shown in the tables above.

For the test phase with non-speech stimuli (figure 5), the actress was shown on a split-screen clapping hands in different tempos. The tempos were 157%, 101%, 63% and 49% of the original recording tempo. The audio was manipulated to 101% of the original recording tempo, thus synchronized with one of the images shown.

The location of the videos on the split screen is given in the tables below.

Baseline for rhythm (clapping movements with no sound)

| | |
|---|---|
| **101%** | 49% |
| 63% | 157% |

Test phase for clapping rhythm. The spatial organization of the videos is the same as during baseline

Figure 5. Hand-clapping in four different tempos: 101% (UL), 49% (UR), 63% (LL), and 157% (LR) of the original tempo. The sound of hands clapping was synchronized with the target image (UL).

except that now the sound track corresponding to the 101% speed plays the clapping sounds.

Baseline for coupling of clapping sounds with visual displays of [by] utterances. The display shows four images of the actress uttering [by] at different speeds. During the baseline the soundtrack is silent.

| | |
|---|---|
| 63% | 49% |
| **101%** | 157% |

The video films for this test phase were identical to those of the baseline but now the sound track played a clapping sound synchronized with the articulatory movements shown on the lower left quadrant.

Half of the subjects were exposed to the speech sound part of the experiment first, followed by the hand-clapping part. The rest of the subjects were exposed to the two parts in reversed order.

## *6.3 Material*

The equipment used for tracking the infant's eye movements was Tobii 1750 eye-tracker integrated with a 17" TFT monitor. The system uses low intensity infrared light to create a static reference frame on the spherical surface of the eye and derives a gaze vector from the relative position of the pupil within that frame. The system performs gaze measurements 50 times per second and with a nominal accuracy of $0.5^{o}$. The eye-tracking data generated by the ClearView 2.2.0 software package that comes with the system was subsequently analyzed using Matematica 5.1 and SPSS 13.0.

## *6.4 Procedure*

The experiments were carried out in a dimly lit studio where most of the light came from the screen connected to the eye-tracking system. The brightness of the display on this screen was enough to draw the infant's attention towards the stimuli being presented. The infant sat in front of the screen at a distance of approximately 60 cm. The parent sat in the studio slightly behind and outside the infant's visual field and listened to music played through sound-isolating head-phones equipped with active noise reduction. Before recording the gaze the system was calibrated using the infant's fixations on special purpose calibration points that were displayed on an otherwise empty screen. The calibration procedure was typically carried out in less than one minute.

# 7. Results

An example of the infants' responses during one of the speech sound conditions is displayed in figure 6. Each panel corresponds to a quadrant on the test screen. The condition illustrated in figure 6 refers to the 20 s video sequence during which the infants head the vowel [a] for the first time in the session. The video corresponding to the sound track was displayed on the lower left corner (LL) of the screen. The curves in each of the panels show the percentage of infants who, at a given time throughout the 20 s of that test phase, were looking at the quadrant represented by the panel. The curves indicate a looking preference towards the upper quadrants, with a slight dominance for the upper right quadrant, displaying the articulation of [ba] syllables. The upper left quadrant, receiving the next highest percentage of looking time through this test phase, displayed the articulation of [by]. The correct visual target was in fact displayed on the lower left quadrant in this test phase and appears to have in fact received the lowest average percentage of looking time.

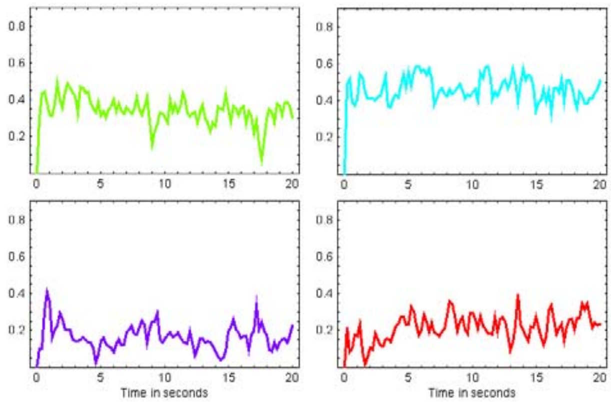


Figure 6. Percentage of infants looking towards each of the quadrants on the screen as a function of time. A running time-window of 200 ms was used. First presentation of [a]. The target image was placed on the lower left (LL) quadrant in this case.

To obtain the net individual gains in looking time towards each of the quadrants shown during the test phases, a repeated measures analysis of variance was performed using the looking times towards a given quadrant during the baseline and the test phase. The

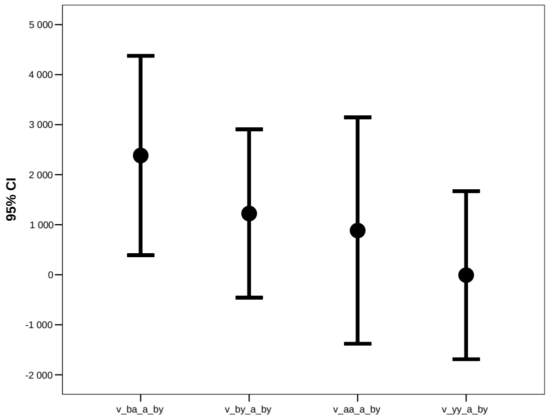


Figure 7. Average looking time gains (in ms) and respective 95% confidence intervals for looking towards the different visual displays of utterances while listening to [by]. From left to right, visual [ba], visual [by], visual [a] and visual [y].

results, using film order as between subject's order, indicated a significant gain for the upper left quadrant displaying [ba], when [by] was heard ($F(1,34)=5.243$, $p<0.028$). There was no significant interaction with film order but film order was a significant between subjects effect ($F(1,34)=4.303$, $p<0.046$). These results suggest thus that the infants matched the visual image of [ba] with the sound of [by], although the group of infants who started the session seeing the clapping sequences performed not as well as the group seeing first the speech stimuli. Another significant gain in looking behaviour was observed for seeing [ba] and when listening to [a] ($F(1,34)=6.196$, $p<0.018$). No significant interaction with film order or significant effect of film order was observed in this case.

The analysis of the looking behaviour during the baseline phases indicates that the infants tended to look most of the time towards the upper quadrants. To compensate for this bias, the results from were sorted in terms of video materials being displayed, rather than the quadrants on which they appeared. Thus, the total looking time towards visual [ba] while listening to [by], for instance, was computed by adding the gain in looking time towards the upper left quadrant during the first test phase where [by] was heard, with the gain in looking time towards the lower left quadrant during the other test phase during which [by] was played. The results from this type of analysis are shown in figures 7 and 8. ANOVA models using the individual subject's looking times towards each of the visual displays shown in figure 7 revealed a within-subjects significant linear trend in looking behaviour towards [ba], [by], [a] and [y] ($F(1,35)=7.235$, $p<0.011$). A very significant within-subjects linear trend was also found for the pattern displayed in figure 8 ($F(1,35)=27.507$, $p<0.0005$).

## *7.1 Non-speech sounds*

The same type of analysis was carried out for the video films involving clapping sounds. In this case all the four quadrants displayed the same type of action but the action was performed with different repetition rates. The results from the infant's matching between clapping sounds and the videos showing the actress clapping at different rates are shown in figure 9. The gain in looking time is greatest for the upper left quadrant, which also is the quadrant showing the clapping movements in synchrony with the sound.

In the other situation involving clapping sounds the videos displayed the actress rhythmically uttering [by]. The results in this case did not show maximum looking time gain towards the lower left quadrant containing the utterances synchronized with the clapping sounds, as illustrated in figure 10. Instead, the maximum looking time was towards the lower right quadrant. However,

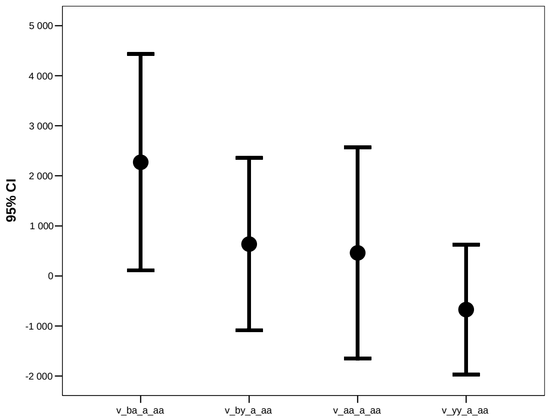


Figure 8. Average looking time gains (in ms) and 95% confidence intervals for looking towards the different visual displays of utterances while listening to [a]. From left to right, visual [ba], visual [by], visual [a] and visual [y].

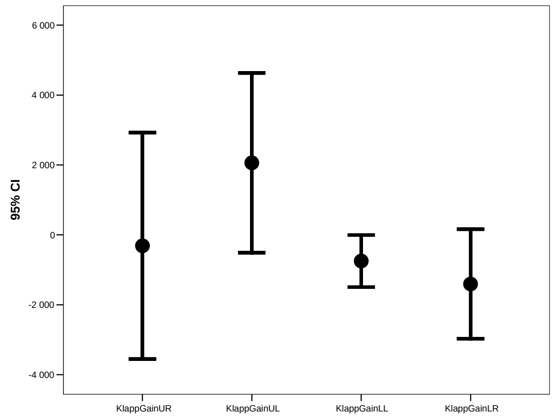


Figure 9. Average looking time gains (in ms) and 95% confidence intervals for looking towards rhythmically produced clap sounds, as illustrated in figure 5. The order of the displayed quadrants is upper right, upper left, lower left and lower right.

when the looking behaviour is organized as a function of the repetition tempo, a pattern of increasing looking times for increasing frequency in the repetition of the articulatory movements emerges. This is significant linear trend ($F(1,35)=9.365$, $p<0.004$).

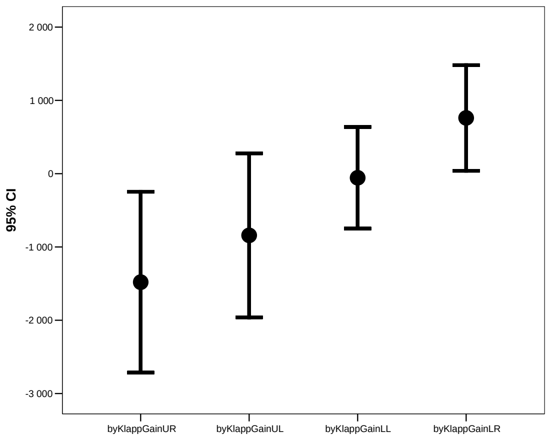


Figure 10. Average looking time gains (in ms) and 95% confidence intervals for looking towards [by] while hearing clapping sounds.

# 8. Discussion

The results of these experiments do not support the strong notion that 8-month-old infants might be able to establish phonetically relevant correspondences between speech sounds and their underlying articulatory movements. Indeed, rather than looking at the quadrants displaying the articulatory movements associated with the speech sounds, the infants' preferences seemed to follow the salience of the articulatory displays on the quadrants (ba>by>a>y). But this was not because they were unable to detect synchrony in general terms. As demonstrated by the tests with clapping sounds, the infants were able to pick up the correct audio-visual synchrony when clap sounds and images were present but they appear to treat speech sounds (or the articulatory movements associated with speech sounds) in a different way than non-speech sounds. In fact, the infants failed to detect synchrony between the non-speech sound and the synchronic articulatory movements. They looked instead longer towards the video film showing the most rapid alternations between closed and open lips, a response that is in line with the results from the "speech sound part" of the experiment.

# 9. Conclusion

Taken together, the two experiments reported here suggest that 8 month-old infants may be using unspecific associative functions to pick up relevant linguistic information on the basis of multi-sensory regularities available in their immediate linguistic environment. If this is true, the infant's success in acquiring the relevant linguistic functions is in line with ETLA and may be more dependent on the structure of its linguistic ambient than on the unfolding of a language acquisition program. In addition, from the point of view of epigenetic robotics this may be a general productive approach, worth to pursue (Dominey & Boucher, 2004). Indeed, given the repetitive characteristics of speech directed to infants (IDS) about 3 months of age, ETLA suggests that it may be possible to derive meaning from the recurrent co-occurrences of auditory and other sensory information representing the infant's immediate linguistic environment. In our MILLE-project, we are currently making efforts to model the early stages of language acquisition exploring the acoustic regularities available in repetitive IDS.

# Acknowledgements

Research carried out with grants from the Swedish Research Council and the Bank of Sweden Tercentenary Foundation.

# References


Bahrick, L. E. (2004). The development of perception in a multimodal environment. In G.Bremner & A. Slater (Eds.), *Theories of Infant Development* (pp. 90-120). New York: Blackwell Publishing.

Dominey, P. F. & Boucher, J.-D. (2004). Developmental stages of perception and language acquisition in a perceptually grounded robot. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, & C. Balkenius (Eds.), (pp. 43-50). Genova, Italy.

Fernald, A., Swingley, D., & Pinto, J. P. (2001). When half a word is enough: infants can recognize spoken words using partial phonetic information. *Child Dev., 72,* 1003-1015.

Gogate, L. J. & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *J.Exp.Child Psychol., 69,* 133-149.

Gustavsson, L., Sundberg, U., Klintfors, E., Marklund, E., Lagerkvist, L., & La (2004). Integration of audi-visual information in 8-month-old infants. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, & C. Balkenius (Eds.), (pp. 143-144). Genova, Italy.

Kuhl, P. K. & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science, 218,* 1138-1141.

Lacerda, F. (2003). Phonolgoy: An emergent consequence of memory constraints and sensory input. *Reading and Writing: An Interdisciplinary Journal,* ***16,*** 41-59.

Lacerda, F., Klintfors, E., Gustavsson, L., Lagerkvist, L., Marklund, E., & Sundberg, U. (2004a). Ecological Theory of Language Acquisition. In Genova: Epirob 2004.

Lacerda, F., Marklund, E., Lagerkvist, L., Gustavsson, L., Klintfors, E., & Sundberg, U. (2004b). On the linguistic implications of context-bound adult-infant interactions. In Genova: Epirob 204.

Lacerda, F. & Sundberg, U. (1996). Linguistic strategies in the first 6-months of life. In (pp. 2574). JASA.

Lidz, J., Gleitman, H., & Gleitman, L. (2003). Understanding how input matters: verb learning and the footprint of universal grammar. *Cognition, 87,* 151-178.

Minsky, M. (1985). Communication with Alien Intelligence. In Edward Regis (Ed.), *Extraterrestrials: Science and Alien Intelligence* ( Cambridge, MA: Cambridge University Press.

Nowak, M. A., Plotkin, J. B., & Jansen, V. A. A. (2000). The evolution of syntactic communication. *Nature, 404,* 495-498.

Prince, C., Hollich, G., Helder, N., Mislivec, E., Reddy, A., Salunke, S. et al. (2004). Taking Synchrony Seriously: A Perceptual-Level Model of Infant Synchrony Detection. In L. Berthouze, H. Kozima, C. Prince, G. Sandini, G. Stojanov, G. Metta, & C. Balkenius (Eds.), (pp. 89-96). Genova, Italy.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition, 71,* 73-108.

# Ongoing Emergence:
# A Core Concept in Epigenetic Robotics

Christopher G. Prince, Nathan A. Helder, & George J. Hollich[*]

Department of Computer Science
University of Minnesota Duluth
Duluth, MN 55812 USA
chris@cprince.com, nhelder@nerp.net

[*]Purdue University
Department of Psychological Sciences
West Lafayette, IN 47907 USA
ghollich@psych.purdue.edu

## Abstract

We propose *ongoing emergence* as a core concept in epigenetic robotics. Ongoing emergence refers to the continuous development and integration of new skills and is exhibited when six criteria are satisfied: (1) continuous skill acquisition, (2) incorporation of new skills with existing skills, (3) autonomous development of values and goals, (4) bootstrapping of initial skills, (5) stability of skills, and (6) reproducibility. In this paper we: (a) provide a conceptual synthesis of ongoing emergence based on previous theorizing, (b) review current research in epigenetic robotics in light of ongoing emergence, (c) provide prototypical examples of ongoing emergence from infant development, and (d) outline computational issues relevant to creating robots exhibiting ongoing emergence.

## 1. Introduction

Epigenetic robotics is a new field that focuses on modeling cognitive development and creating robots that show autonomous mental development (Lungarella, Metta, Pfeifer, & Sandini, 2003; Weng, McClelland, Pentland, Sporns, Stockman, Sur, & Thelen, 2001). For example, robots have been implemented that generate visual discrimination behavior using large-scale neural networks (Seth, McKinstry, Edelman, & Krichmar, 2004), that model early infant-caregiver interaction using behavioral rules (Breazeal & Scassellati, 2000), and that explore the knowledge needed by infants to succeed in perceptual object permanence experiments (Chen & Weng, 2004; Lovett & Scassellati, 2004; see also: Schlesinger & Casey, 2003). Given these and other diverse contributions to this new field (for a review, see Lungarella et al., 2003) it seems an opportune time to synthesize a few core concepts from this corpus of research.

In this paper, we distill one such core concept, *ongoing emergence*, which refers to the continuous development and integration of new skills. An agent exhibiting ongoing emergence, in a motivationally autonomous manner, will continue develop and refine its skills across development. This vision of open-ended development is evident in recent work. For example, in efforts to "allow a mobile robot to incrementally progress through levels of increasingly sophisticated behavior" (p. 1, ms., Blank, Kumar, Meeden, & Marshall, 2005), in efforts to build robots that exhibit "new behavior, which in turn, becomes a precursor for successive stages of development" (p. 27, Grupen, 2003), and in efforts to achieve robots exhibiting a "successive emergence of behaviors in a developmental progression of increasing processing power and complexity" (p. 1, ms., Dominey & Boucher, 2005). Unfortunately, while humans clearly show such long-term progressions, epigenetic robots as yet do not—they are typically designed to achieve particular behaviors or to learn specific tasks.

To escape this impasse, we propose a theoretical framework for achieving ongoing emergence. To this end, in Section 2 we review previous theoretical conceptions regarding ongoing emergence and synthesize the current state of the art in terms of six criteria. Section 3 considers how current examples of robotic systems fare with respect to these criteria for ongoing emergence. In Section 4, we look to infant developmental research for examples of ongoing emergence. Section 5 outlines some computational issues for designing robots that exhibit ongoing emergence. We close with a discussion.

## 2. Conceptual Synthesis

### *2.1. Background*

Blank et al. (2005) discuss the possibility that a robot can use a *developmental algorithm* to learn, via a process of self-exploration, its repertoire of behaviors and mental capabilities, instead of being preprogrammed with "the capabilities of a human body and human concepts" (p. 2, ms). Robots are proposed to discover even the most primitive behaviors through a process of exploration.

A possible benefit of providing such a developmental algorithm is avoiding specification of task-goals for the robot. Instead, "it is the goal of developmental robotics to explore the range of tasks that can be learned (or grown) by a robot, given a specific developmental algorithm and a control architecture" (p. 2, ms). These authors consider three mechanisms to be essential to developmental algorithms: abstraction, anticipation, and self-motivation. Abstractions are seen as necessary to focus the robot's attention on relevant environmental features, given the "constant stream of perceptual information" (p. 2, ms.). Anticipations enable the robot to predict environmental change to "go beyond simple reflexive behavior to purposeful behavior" (p. 3, ms.). And self-motivation "push[es] the system toward further abstractions and more complex anticipations" (p. 3, ms.). It is thought that a developmental algorithm incorporating these three mechanisms could be successively applied to move an agent from a discovery of initial behaviors ("reflexes") to more complex behaviors.

Weng (Weng, 2004; Weng et al., 2001) also emphasizes the need for robots to autonomously generate their own task-specific representations in order to cope with dynamic, unknown, or uncontrolled environments. "A developmental program for robots must be able to generate automatically representations for unknown knowledge and skills" (Weng et al., 2001) so as to adapt to these environmental variations. An agent with the capacity to construct its own representations has the potential of understanding these representations. "Without understanding, an agent is not able to select rules when new situations arise, e.g. in uncontrolled environments" (p. 205, Weng, 2004). These processes are viewed as open-ended and cumulative. "A robot cannot learn complex skills successfully without first learning necessary simpler skills, e.g., without learning how to hold a pen, the robot will not be able to learn how to write" (Weng et al., 2001).

Grupen (2003) is similarly concerned with enabling robots to solve "open tasks in unstructured environments" (p. 2, ms.). The approach he advocates is to use "developmental processes [that] construct increasingly complex mental representations from a sequence of tractable incremental learning tasks" (p. 1, ms.). He proposes "computational mechanisms whereby a robot can acquire hierarchies of physical schemata" (Grupen, 2003, p. 1, ms.). Physical schemata provide parameterized, and in that sense reusable, sensorimotor control knowledge.

Dominey and Boucher (2005) model linguistic grammar acquisition, based on visual and auditory pre-processing of sensory inputs and connectionist models. The authors use the developmental theory of Mandler (1999), who "suggested that the infant begins to construct meaning from … scene[s] based on the extraction of perceptual primitives. From simple representations such as contact, support, and attachment … the infant [may] construct progressively more elaborate representations of visuospatial meaning" (p. 244, Dominey & Boucher, 2005).

## 2.2. Synthesis

From this earlier thinking, we wish to synthesize a picture of what we refer to as ongoing emergence. We propose six defining criteria for ongoing emergence (see Table 1). Our first two criteria are: (1) An agent creates new skills by utilizing its current environmental resources, internal state, physical resources, and by integrating current skills from the agent's repertoire, and (2) These new skills are incorporated into the agent's existing skill repertoire and form the basis from which further development can proceed. By "skills" we include overt behaviors, perceptual abilities, and internal representational schemes.

These first two criteria express the notion that when we view agents as developing systems, with certain skills in their repertoire, they have the potential to develop related skills. For example, under this view a developmental robot that can learn to kick a ball might then later develop skills for playing soccer. Ongoing emergence thus has the property of *developmental systematicity*[1]. In developmental systematicity if an agent demonstrates skill $a\boldsymbol{R}b$, then we also expect competence with directly related skills, $b\boldsymbol{R}a$ (i.e., systematicity; Fodor & Pylyshyn, 1988). Furthermore, we expect the emergence of developmentally related skills such as $f(a)$ and $g(a\boldsymbol{R}b)$, where $f(x)$ and $g(y)$ are developmental processes producing emergent skills in the agent's repertoire over time. This process would in part be based on earlier skills (e.g., $x$ and $y$ in $f(x)$ and $g(y)$ above). For example, if a robot exhibits a range of object tracking behaviors ($a\boldsymbol{R}b$, $b\boldsymbol{R}a$) through the composition of blob tracking skill ($a$) and motion finding skill ($b$), and the robot is a developing agent, we would have further, developmental expectations about its future behaviors such as facial tracking and gaze following (e.g., $f(a)$, $g(a\boldsymbol{R}b)$).

Another central notion in the work described in Section 2.1 is that of autonomy: avoidance of specification of task goals, autonomous generation of task-specific representations, and the ability to solve open tasks. We include this as a third criterion for ongoing emergence: (3) An agent that exhibits ongoing emergence autonomously develops adaptive skills on the basis of having its own values (e.g., see Sporns, 2005; Sporns & Alexander, 2002) and goals, with these values and goals being developed by the system in a manner similar to its skills. If an agent develops its own values and goals, it can use these for self-supervision and to determine the tasks that need to be solved. In brief, the agent needs some way to evaluate its own behaviors, and determine when a particular skill is useful[2]. This is true in both the short and long term. For example, in the short-term, a robotic agent might tradeoff energy output for the gain of information, while long-term goals might include improving communication amongst the robot's cohorts.

To these initial three criteria for ongoing emergence we add three additional criteria: (4) bootstrapping (when the system starts, some skills rapidly become available), (5) stability (skills persist over an interval of time), and (6) reproducibility (the same system started in similar initial states and in similar environments also displays

[1] We introduce the concept of developmental systematicity to avoid viewing behavior generation an infinite domain. This is analogous to the way that Fodor & Pylyshyn (1988) introduced systematicity to avoid viewing language generation as an infinite domain.

[2] We refrain from adopting the idea that skills that emerge in development should necessarily be more complex (i.e., be more powerful in some sense) than prior skills. From our view, this criterion is too strong for several reasons. First, strictly increasing adaptation is violated in some instances of child development (e.g., the "U" shaped curves of child performance on various tasks over time; see Siegler, 2004). Second, a view of strictly increasing complexity of skills does not allow for escape ("detours") from local maxima, where behavior needs to get worse before it can get better. Third, strictly increasing skill complexity may remove the possibility of discovering simpler means to achieve the same (or similar) ends as existing skills—as in evolution, "different" is sometimes at least as good as "better." Relatedly, a strict view of building complexity does not seem to allow for the loss (e.g., forgetting) of some skills over time.

similar ongoing emergence). We include bootstrapping as a criterion for ongoing emergence because it seems inevitable that a robot either needs to have some means of spontaneously developing its own set of initial skills or, more conventionally, will need to have some initial skills pre-programmed prior to its being turned on. While pre-programming of bootstrap skills is not consistent with the concept of skills being developed by the agent itself, we consider this an acceptable practice if for no other reason than keeping the scope of research projects tractable. However, in our view, the preferred method to establish bootstrapping skills is to represent those skills in the same manner as later emerging skills such that both the bootstrapping and developed skills comprise a uniform part of the agent's skill repertoire.

Stability of skills is in part a practical matter: in order for a skill to be measured (i.e., by researchers), it must exist for a measurable duration. In terms of the robot, however, stability may be more than merely a practical matter in that, in order for ongoing emergence to be achieved, a certain degree of skill stability over time will likely prove necessary. If the behaviors exhibited by the robot are merely transient then those behaviors may not contribute to the basis for the acquisition of new skills.

The reproducibility criterion asks the question: Under what starting-state and environmental conditions does a given developmental algorithm produce an ongoing emergence of behavior? We presume that if a developmental algorithm is well-understood, then the starting-state and conditions under which it produces ongoing emergence will also be well-understood. These conditions may be limited (e.g., to specific values of initial variables), but once known can reproducibly generate ongoing emergence.

# 3. Current Research & Ongoing Emergence

In this section, we review examples of epigenetic robotic research (see Table 2) in light of our criteria for ongoing emergence. Our selection of these particular papers is not intended to reflect some *a prior* sense that they have achieved ongoing emergence. Rather they simply reflect our subjective impression of good illustrative examples of research in this area.

Several lines of research satisfy Criterion 1 (new skill acquisition). For example, a swinging behavior emerges in the robot of Berthouze and Lungarella (2004), and the skill of tracking a face view to an object emerges in the robot of Nagai et al. (2003). To varied extents, some research has also satisfied Criterions 3 through 6. Criterion 3 (autonomy of goals and values) is satisfied to some degree by the robot of Seth et al. (2004), and also the work of Kaplan and Oudeyer (2003). Seth et al. (2004) utilize a value system in the Darwin VIII robot to signal the occurrence of salient sensory events. Initially, Darwin VIII's value system was activated by sounds detected by the robot's auditory system, but through learning became activated by particular visual stimulus attributes. Criterion 4 (bootstrapping) is satisfied by the Dominguez and Jacobs (2003) system in that the system uses progressive changes in visual acuity to increase its binocular disparity sensitivity. Criterion 5 (stability) appears satisfied, for example, by the Lungarella and Berthouze (2002) system in that the robots' swinging behavior reaches stable states. Criterion 6 (reproducibility) is satisfied by studies which replicated their robots' behavior, perhaps under varied conditions. For example, Chen and Weng's (2004) experiments were replicated with 12 separate robot "subjects" (the same robot and algorithms, but with different environmental conditions).

To give a more extended example of how these criteria can be applied, we consider the work of Nagai et al. (2003). These authors modeled joint visual attention behavior using a robot. Joint attention occurs when individuals both look at the same object, and may include knowledge of shared attentional states (e.g., Carpenter, Nagell, & Tomasello, 1998). The Nagai et al. (2003) robot learned to track the face view of a person to the object the person was looking at. Learning started with the robot (a) knowing how to visually detect faces and salient objects, (b) knowing to switch its gaze from a face to an object when both the object and face were in its field of view, and (c) having a predefined transition function (a sigmoidal) to switch from how salient objects were found—either directly in its visual field, or indirectly by first looking at a face[3]. Initially the robot did not know how much to turn its head based on a particular view of a face to find the object that the person was looking at, and learning acquired this skill. The transition function enabled this skill to be gradually applied.

The Nagai et al. (2003) robot has some behaviors that are programmed into the system (e.g., bootstrapping, Criterion 4). A new behavior is constructed from the initial behaviors (e.g., visually detecting faces and objects) and environmental interaction—i.e., the robot learns to track faces to the objects that they are looking at (Criterion 1 is satisfied). However, once the new behavior has emerged, there is no further potential for development. That is, the new skill was not incorporated into the system in such a way that it contributed to the basis for further skill development. Thus, Criterion 2 is not satisfied. Reproducibility (Criterion 6) was demonstrated in the system by conducting experimental runs with 1, 3, 5, or 10 objects in which the emergent behavior was maintained. In summary, while some of the criteria for ongoing emergence are satisfied, the behavior of the Nagai et al. (2003) robot seems best classified as demonstrating emergence and not ongoing emergence.

Notably absent in this review of current work is full evidence for Criterion 2 (incorporation of new skills with existing skills so that those new skills can be used as part of the basis for further development). We have yet to find examples of robots exhibiting this ability (but we hope to be corrected on this point!). This leads us to view current

[3] The use of this sigmoidal seems unnecessary, but in our view should not be viewed as a shortcoming of this research. In principle, the authors could have used a method of self-supervised learning to transition between modes: when the robot was sufficiently able to predict the amount of head turn required for accurately turning to face an object, it could have then begun utilizing its self-generated head turn.

examples of epigenetic robots (e.g., see Table 2) as demonstrating emergence, but not ongoing emergence.

# 4. Human Infant Developmental Examples

In contrast to the state of the art in epigenetic robotics, human infants clearly exhibit ongoing emergence. Development is an unending process that continually produces new skills by making use of currently available skills, environmental conditions, and other resources. In this section, we provide prototypical examples of ongoing emergence in infants from three developmental areas: walking, language, and visual object skills.

## *4.1 Emergence of Walking*

As any one-year-old would acknowledge, walking is more difficult than it may first appear. To properly walk, children must achieve the right mix of balance, head control, and coordinated oscillation of the limbs that have thousands of muscle fibers and billions of nerves as well as their own length, mass and transitory inertia. The degrees of freedom for the task are enormous (Bernstein, 1967).

Further complicating this process, children grow physically. They begin life top-heavy, which makes stabilizing this system all the more difficult. Their growth is also erratic and dramatic—children can go up to 63 days with no measurable change in height and then suddenly grow up to 2.5cm in a single night (Lampl, Veldhuis, & Johnson, 1992). As if this wasn't difficult enough, children must learn to navigate different slopes and uneven terrain, perhaps while carrying objects (Adolph & Avolio, 2000). Yet somehow nearly all children learn to walk, and continue to walk, despite the complexity of achieving the right mix of skills, changes in body morphology, and varying situations.

Current theory (Thelen, 1995) views this process as a dynamic self-organizing system in which integration of diverse skills plays a key role. Because the world, the task, and even children's bodies are constantly changing, each component is constantly being weighted differently, as dictated by the interaction of the nervous system and the environment. For example, while all infants posses a stepping and a kicking reflex at birth, the stepping reflex disappears after a few months. Why? In short, babies don't have the strength to keep up this reflex as they grow heavier—even though the nervous system is still sending the signals. Stepping and, by extension, walking must wait for stronger muscles to grow before infants can take their first steps. If one makes the task easier, by supporting the infants (on a treadmill or underwater), even newborns can walk (Thelen & Fisher, 1982). In contrast, if one makes the task harder by placing weights on older infants, their walking again approximates that of younger infants (Thelen & Fisher, 1982). Thus, it is the dynamic interaction between current skills, the state of the system, and the environment that allows for walking behavior to self-organize into coherent patterns across changes in morphology and task. This illustrates Criterion 1 for ongoing emergence—namely that skills are created through the integration of resources including environment and existing skills.

## *4.2 Emergence of Language*

While purely physical skills like walking show ongoing emergence, skills that are more cognitive also require the use and integration of multiple developing skills. For example, word learning can be seen as the product of social skills (e.g., sensitivity to eye gaze), domain-general skills (e.g., sensitivity to statistical structure such as invariances), and linguistic skills (e.g., a bias toward labeling objects based on shape).

Just as in walking, the weight placed on each of these skills likely changes with time and situation. In the beginning, infants may depend on a range of perceptual biases and statistical relations to establish the meaning of each new word (Hollich, Hirsh-Pasek, & Golinkoff, 2000). However, as more words are learned, children use their knowledge of known words to help them learn additional words. This illustrates another property (Criterion 2) of systems exhibiting ongoing emergence: incorporation of new skills into a skill repertoire. For example, Smith (1999) provides evidence that infants may notice how particular types of words get extended (e.g., nouns are generalized, a.k.a. extended, to different objects on the basis of shape). Infants then use this knowledge to guide their own extensions of novel words. When told a U-shaped object is a "dax," infants will spontaneously extend that word to other U-shaped objects. Even so the system is flexible—infants will not extend a word based on shape if the object happens to have eyes (or even shoes), suggesting that children have noticed that living creatures can often change their overall shape in ways that static objects do not.

Related to the use of multiple skills, the more skills that an agent can bring to bear, the more fault-resistant and flexible the system. Loss of one skill does not cause the system to fail entirely, and the interaction among skills insures that children can successfully acquire a language under extremely impoverished conditions. For example, even deaf children growing up in an area without exposure to any fully formed language will create their own language (Sengas, 1995). With both biological and robotic systems, more pathways to success imply greater adaptivity and increased likelihood of organism survival. With regular upheaval at the neurological and muscular levels, it is no wonder that developmental architectures are massively fault tolerant with multiple, redundant skills. Thus, the self-organization of new skills combined with an increasing skill repertoire can lead to a process of ongoing emergence.

## *4.3. Object Skill Developments*

At the same time that human infants are developing the walking and language acquisition skills discussed above, they also show an ongoing emergence of physical and mental capabilities related to visual objects. Starting from birth, infants are able to extract information about object size and shape (Slater & Morison, 1985), remember objects over time (Slater, Morison, & Rose, 1982), and perceive similarities and differences between visual stimuli (Slater, Morison, & Rose, 1984). Newborn infants are also able to track a moving target with eye and head movements (albeit in a jerky fashion, e.g., Aslin, 1981; von Hofsten 1982), and can recognize the constancy of an object's identity across transformations in orientation and movement (Slater, Morison, Town & Rose, 1985).

While constituting a perhaps surprisingly robust set of initial skills, developing and incorporating these skills into more complex behaviors takes time. For example, it is not until 4 months of age that an infant's muscular control and object understanding have matured to the point of allowing an infant to successfully reach for and grasp an object (e.g., von Hofsten, 1989). Also at 4 months, infants begin to perceive (measured via looking-time) a partially occluded object as a single unified object (Kellman & Spelke, 1983; Johnson & Nanez, 1995). However, it is not until about 6 months of age that infants combine their object tracking skills, their understanding of object unity, and their reaching skills to reach for an object that has been partially obscured from view by an occluding object (von Hofsten & Lindhagen, 1979; Shinskey & Munakata, 2001).

Also in the realm of visual-object skills is object permanence, which relates to the child's understanding that an object continues to exist even when the object cannot be seen. It has been shown that 3.5-month-old infants show recognition of an impossible object event (i.e., a violation of object permanence), such as a drawbridge closing despite a solid object appearing to have been blocking its path (Baillargeon, 1987, 1993, 1995). However, this sort of "perceptual object permanence" is not manifested as a behavior indicating an understanding of physical (i.e., more conventional) object permanence until much later, when 8- to 10-month-old infants will begin to search for an object that has been hidden from view (Piaget, 1954). Still, infant searching behavior at this age is not free of difficulties and is subject to the "A-not-B error" (the infant searches for a hidden object at location A when the object was initially uncovered at location A but subsequently hidden at location B). Infants perseverate in this error until roughly 12 months of age (at which time infants will correctly search for the hidden object at location B; e.g., see Wellman, Cross, & Bartsch, 1986; Newcombe & Huttenlocher, 2000).

In developing from initial skills of being able to identify and track objects (birth), to perceptually distinguishing impossible object events (3.5 months), to being able to maintain perception of object unity despite an occlusion (4 months), to successfully reaching for an object (4 months), to successfully reaching for an object despite an occlusion (6 months), to searching for a hidden object (8-10 months), to searching for a hidden object without displaying the A-not-B error (12 months), infants demonstrate an ongoing emergence of behavior. Changes occurring in the visual, conceptual, and motor systems of the infants interact to produce unique, observable behaviors at multiple points along the developmental path of these visual object skills, with each developed skill being incorporated and providing a contributing factor to the emergence of subsequent skills.

# 5. Designing For Ongoing Emergence

Past a theoretical understanding of ongoing emergence, our most burning question was well-expressed by one of the anonymous referees of this paper: How can we design robots so that the behaviors exhibited by the robot continue to be adaptive and open to further development throughout their duration of use (e.g., either as models of infants, or deployed in some industrial environment)? That is, how do we design robots that exhibit ongoing emergence? Our thinking here divides broadly into two possibilities. The first possibility we address is that of designing robots that exhibit ongoing emergence where the bootstrapping components (see Criterion 4) of the system are not generated by ongoing emergence. Effectively, this corresponds to basing the design of the robots on existing research (e.g., the robotic systems in Table 2). The second possibility we address is that of designing robots that exhibit ongoing emergence where the bootstrapping components themselves are generated by processes of ongoing emergence. This corresponds to discovering a different way of approaching the design of the initial components of a robotic system.

## *5.1. Bootstrapping Ongoing Emergence **Without** Primitive Ongoing Emergence*

Ongoing emergence in humans results in part from the dynamic integration of multiple skills with the environment (i.e., Criterion 1). One way to achieve an analog of this in robots may be to add an integration layer on top of an existing system or systems (see Table 2), providing soft-assembly of the component skills. For example, we might combine robotic behaviors across several systems, such as the perceptual object permanence behavior of Cheng and Weng (2004), the joint attention of Nagai, et al. (2003) and the social skills of Breazeal and Scassellati (2000). For this integration layer to satisfy Criterion 1, it would be appropriate for these skills, in interaction with the environment, to produce new, adaptive, emergent skill(s). For example, given the integration of the above three prior research projects, the integrated system might express surprise towards a caregiver when an object permanence situation was violated.

An approach that might be useful for this integration layer was given by Cheng, Nagakubo, and Kuniyoshi (2001). These authors proposed an integration mechanism to combine components in a humanoid robotic system, involving integrating the results of various component mechanism, which themselves show adaptation over time. Combining components involves weighting the components for their relative contributions, and such contributions may vary according to factors such as learning and context. The authors use a sensory-cue competition approach to integration, and generating motor outputs. They define the motor output of a robot as the vector $U_i(t)$, expressed by equation [1]. $U_i(t)$

gives the output for motor sub-system *i* (e.g., a head control motor), at time *t*.

$$U_i(t) = \frac{\sum_k \alpha_k(t) a_k(t) v_k(t)}{\sum_k a_k(t)} \quad [1]$$

In equation [1], $v_k(t)$ is a vector giving the current sensory input from sensory subsystem *k* (e.g., a joint angle or a camera) at time *t*, $a_k(t)$ is a measure of the reliability (confidence) of sensory subsystem *k* (a scalar), and $\alpha_k(t)$, defines the strength (e.g., priority) of a particular sensory input (also a scalar).

In a perceptual context, the weighted component integration (or *democratic integration*) algorithm forwarded by Triesch and von der Malsburg (2001) offers similar ideas, and presents a more detailed investigation of the integration concept than Cheng et al. (2001). In Triesch and von der Malsburg (2001), a group of perceptual components such as motion, color, and shape detection are adapted both in terms of the weighting of the components and in terms of prototypes for the perceptual components.

Unfortunately, the integration mechanisms proposed by both Cheng et al. (2001) and Triesch and von der Malsburg (2001) do not focus on or provide specific means of incorporating skills that result from the process of integration—leaving Criterion 2 unsatisfied. Two additional computational mechanisms would seem needed past Cheng et al. (2001) and Triesch and von der Malsburg (2001) in order to provide skill incorporation. First, the system needs a (at least implicit) means of determining that a soft-assembled skill is re-occurring. That is, the system needs a way to determine when that skill should be considered "stable" (Criterion 5). Second, once stable, the skill needs to be represented in a manner similar to the existing skills. This last part, at least in terms of this present scenario seems particularly difficult. We have been working from the design premise of utilizing current results from epigenetic robotics to form the bootstrapping components of a robot, intended to display ongoing emergence. But, in this case, there is no particular means to add to this static collection of bootstrapping skills. Presumably, a computational mechanism would be needed to learn the aspects of the new, now-stable soft-assembled skill. This new skill would, after this learning, be part of the repertoire of the system and with the other skills would form the basis for further development (i.e., it would then be termed *incorporated*; Criterion 2).

### *5.2. Bootstrapping Ongoing Emergence* ***With*** *Primitive Ongoing Emergence*

A possible limitation of adopting the strategy proposed above is that one may miss common underlying mechanisms that helped create the individual skills in the first place. That is, in the ideal case, the goal would be to create a robot that exhibits ongoing emergence, where the bootstrapping primitives themselves are emergent. Thus, in this ideal case the bootstrapping primitives are generated by the same processes that underlie subsequent skill development.

This presents a rather different problem than in Section 5.1. On the one hand, a robot that has a pre-programmed set of behaviors can presumably exhibit those behaviors (e.g., in a sequence, or through a blending of behaviors, such as shown in Breazeal, Buchsbaum, Gray, Gatenby, & Blumberg, 2005), but is in need of mechanisms to incorporate stabilized soft-assembled behaviors into its skill repertoire. On the other hand, a robot without a pre-programmed set of behaviors, in addition to needing mechanisms to provide ongoing emergence itself, is in need of an initial set of skills—it needs initial means of perceiving, representing, and behaving.

One conceptual way that such initial—and emergent—skills might be created is through self-exploration. A number of authors in epigenetic robotics have suggested the need for some form of "self" in these robotic systems. For example, Weng (2004) proposes that developing robots must be SASE—Self-Aware and Self-Effecting agents, Blank et al. (2005) talk about self-motivation and self-organization, and Steels (2004) suggests that robotic agents should self-regulate their build-up of skills and knowledge as a way to increase their rate of learning. In the present context of bootstrapping a developing agent without pre-programmed skills, self-exploration could be used to facilitate differentiation between self and other (e.g., see Michel, Gold, & Scassellati, 2004), which is important because such a developing agent would likely need to figure out what parts of its "environment" are part of the agent (e.g., its own limbs) versus part of the external world. We hypothesize that the basic properties of ongoing emergence (i.e., Criterion 1 through 6) could also provide the basis for these self-other discrimination skills, and hence can provide the means to bootstrap the skills of a developing robot.

## 6. Discussion

The foregoing has been a largely theoretical discussion of ongoing emergence. Ongoing emergence describes, in brief, behavioral growth in humans and (hopefully, in the future) in robots. If we have achieved our goal, this paper will stimulate further theoretical and empirical research towards these ends. We hope that this is but one of many entries to follow in the continuing discussion of behavioral growth in robots. In closing this paper, we want to argue for a relation between ongoing emergence and theorizing in cognitive science, we discuss additional means by which ongoing emergence may be achieved incrementally in epigenetic robotics research, and we close with a view to the future.

In Section 5 we raised a distinction between using pre-programmed initial skills (Section 5.1) and not using pre-programmed initial skills but instead relying purely on the properties of ongoing emergence (Section 5.2). In the pre-programmed initial skills case, we take this to be analogous to Fodor's concept of modularity (Fodor, 1983). We take this position because the amount of interaction between the components will be limited and because the components show limited development. This provides another way to view the Section 5 alternatives: ongoing emergence through separate modules versus ongoing emergence through "modules" that develop.

It seems crucial to establish methodological ways to achieve research progress in ongoing emergence. While we have implicitly offered some ideas to this end in the body of the paper, three further ideas come to mind. First, it seems conceptually possible that ongoing emergence could be exhibited strictly within particular domains. For example, a robot might exhibit ongoing emergence only in its language and communication skills, or only in its object manipulation skills. Second, it also seems conceptually possible that ongoing emergence may be achieved in a primarily perceptual manner. We feel justified in part for this statement by the productivity of psychological methods with infants that have focused largely on the development of perceptual knowledge (e.g., Baillargeon, 1995; Hollich et al., 2000). Third, a potentially useful research step towards a full sense of ongoing emergence may be a linear emergence of a limited number of skills. In this case, a *single* skill would emerge, that skill would then be incorporated into the robot's existing skill repertoire, and then this new pool of skills would be used to develop *one* additional skill.

In closing, we recollect the statements of György Gergely, in his invited address at EpiRob 2003. György suggested that "recent research in epigenetic robotics has been strongly preoccupied with and [has] made significant advances towards modeling the 'lower level' mechanisms and 'bottom–up' processes involved in systems of action perception and production and the ways in which these systems [may be] inherently interrelated" (p. 192, Gergely, 2003). Clearly, with goals including modeling cognitive development, epigenetic robotics should not be limited to modeling 'lower level' mechanisms. But, how do we make progress? In the terms of this paper, we advocate directly tackling the challenge of ongoing emergence, and in particular our Criterion 2 (incorporation of skills) seems in most need of further research. If cognitive skills arise out of ongoing emergence, then if we achieve robots with ongoing emergence, there is a good chance that those robots will have instantiated models of cognitive skills.

## Acknowledgements

The authors thank Lakshmi Gogate, Rich Maclin, and Eric Mislivec for discussions and comments. A discussion with John Weng started this paper. CGP thanks his Spring05 CS5541 class for discussions about some of the papers reviewed here. We thank the EpiRob05 anonymous referees for constructive feedback.

## References

Adolph, K. E., & Avolio, A. M. (2000). Walking infants adapt locomotion to changing body dimensions. *Journal of Experimental Psychology: Human Perception & Performance*, *26*, 1148-1166

Aslin, R. N. (1981). Development of smooth pursuit in human infants. In D. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye movements: Cognition and visual perception*. Hillsdale, NJ: Erlbaum.

Baillargeon, R. (1987). Object permanence in 3 1/2- and 4 1/2-month old infants. *Developmental Psychology*, *23*, 655-664.

Baillargeon, R. (1993). The object concept revisited: New directions in the investigation of infants' physical knowledge. In C. Granrud (Ed.), *Visual Perception and Cognition in Infancy* (pp. 265–315). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Baillargeon, R., (1995). A model of physical reasoning in infancy. In C. Rovee-Collier & L. P. Lipsitt (Eds.) *Advances in Infancy Research, Vol. 9* (pp. 305–371). Norwood NJ: Ablex.

Bernstein, N. (1967). *The Coordination and Regulation of Movements*. London: Pergamo.

Berthouze, L., & Lungarella, M. (2004). Motor skill acquisition under environmental perturbations: On the necessity of alternate freezing and freeing of degrees of freedom. *Adaptive Behavior*, *12*, 47-63.

Blank, D., Kumar, D., Meeden, L., & Marshall, J. (2005). Bringing up robot: Fundamental mechanisms for creating a self-motivated, self-organizing architecture. *Cybernetics and Systems*, *36*, 125-150.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. *Artificial Life*, *11*, 31-62.

Breazeal, C. & Scassellati, B. (2000). Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior*, *8*, 49-74.

Carpenter, M., Nagell, K. & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development,* Serial No. 255, 63 (4).

Chen, Y. & Weng, J. (2004). Developmental learning: A case study in understanding "object permanence." In *Proceedings of the Fourth International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Lund, Sweden: Lund University Cognitive Studies.

Cheng, G. & Nagakubo, A., & Kuniyoshi, Y. (2001). Continuous humanoid interaction: An integrated perspective—gaining adaptivity, redundancy, flexibility—in one. *Robotics and Autonomous Systems*, *37*, 161-183.

Dominguez, M. & Jacobs, R. A. (2003). Developmental constraints aid the acquisition of binocular disparity sensitivities. *Neural Computation*, *15*, 161–182.

Dominey, P. F. & Boucher, J.-F. (2005). Developmental stages of perception and language acquisition in a perceptually grounded robot. *Cognitive Systems Research*, 6, 243-259.

Fodor, J. A. (1983). *The Modularity of Mind*. Cambridge, MA: MIT Press.

Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture. *Cognition*, *28*, 3-71.

Gergely, G. (2003). What should a robot learn from an infant? Mechanisms of action interpretation and observational learning in infancy. *Connection Science*, *15*, 191–209.

Grupen, R. (2003). A developmental organization for robot behavior. In *Proceedings of the Third International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Lund, Sweden: Lund University Cognitive Studies.

Hofsten, C. von (1982). Eye-hand coordination in newborns. *Developmental Psycholog, 18*, 450-461.

Hofsten, C. von (1989). Transition mechanisms in sensorimotor development. In A. De Ribaupierre (Ed.), *Transition Mechanisms in Child Development: The Longitudinal Perspective* (pp. 233-258). New York, NY: Cambridge University Press.

Hofsten, C. von, & Lindhagen, K. (1979). Observations on the development of reaching for moving objects. *Journal of Experimental Child Psychology, 28*, 158-173.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. (2000). Breaking the language barrier: An emergentist coalition model of word learning. *Monographs of the Society for Research in Child Development*, *65* (3, Serial No. 262).

Johnson, S. P., & Nanez, J. E., Sr. (1995). Young infants' perception of object unity in two-dimensional displays. *Infant Behavior and Development*, *18*, 133-143.

Kaplan, F. & Oudeyer, P-Y. (2003). Motivational principles for visual know-how development. *In Proceedings of the 3rd International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*. Lund University Cognitive Studies.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483-524.

Lampl, M, Veldhuis, J. D., & Johnson, M. L. (1992). Saltation and stasis: a model of human growth. *Science*, *258*, 801-803.

Lovett, A., & Scassellati, B. (2004). Using a robot to reexamine looking time experiments. In *Proceedings of the Third International Conference on Development and Learning*.

Lungarella, M., & Berthouze, L. (2002). On the interplay between morphological, neural and environmental dynamics: A robotic case-study. *Adaptive Behavior*, *10*, 223–241.

Lungarella, M., Metta, G., Pfeifer, R., & Sandini G. (2003). Developmental robotics: A survey. *Connection Science*, *15*, 151-190.

Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003). A constructive model for the development of joint attention. *Connection Science*, *15*, 211-229.

Newcombe, N. & Huttenlocher, J. (2000). *Making Space: The Development of Spatial Representation and Reasoning*. Cambridge, MA: MIT Press.

Mandler, J. (1999). Preverbal representations and language. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and Space* (pp. 365–384). MIT Press.

Michel, P., Gold, K., & Scassellati, B. (2004). Motion-based robotic self-recognition. Presented at the *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Schlesinger, M., & Casey, P. (2003). Where infants look when impossible things happen: Simulating and testing a gaze-direction model. *Connection Science*, *15*, 271-280.

Sengas, A. (1995) The development of Nicaraguan Sign Language via the language acquisition process. In D. MacLaughlin & S. McEwen (Eds.) *Proceedings of The 19th Boston Univeristy Conference on Language Development.* Boston: Casadilla Press.

Siegler, R. S. (2004). U-shaped interest in u-shaped development – and what it means. *Journal of Cognition and Development*, *5*, 1-10.

Seth, A. K., McKinstry, J. L., Edelman, G. M., & Krichmar, J. L. (2004). Visual binding through reentrant connectivity and dynamic synchronization in a brain-based device. *Cerebral Cortex*, *14*, 1185-1199.

Shinskey, J. L., & Munakata, Y. (2001). Detecting transparent barriers: Clear evidence against the means-end deficit account of search failures. *Infancy, 2*, 395-404.

Slater, A., & Morison, V. (1985). Shape constancy and slant perception at birth. *Perception*, *14*, 337-344.

Slater, A., Morison, V., & Rose, D. (1982). Visual memory at birth. *British Journal of Developmental Psychology*, *73*, 5 19-525.

Slater, A., Morison, V., & Rose, D. (1984). Newborn infants' perception of similarities and differences between two- and three-dimensional stimuli. *British Journal of Developmental Psychology*, *2*, 287-294.

Slater, A., Morison, V., Town, C., & Rose, D. (1985). Movement perception and identity constancy in the new-born baby. *British Journal of Developmental Psychology*, *3*, 21 l- 220.

Smith, L. B. (1999). Children's noun learning: How general learning processes make specialized learning mechanisms. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 277-304). Mahwah, NJ: Erlbaum.

Sporns, O., & Alexander, W .H. (2002). Neuromodulation and plasticity in an autonomous robot. *Neural Networks*, *15*, 761-774.

Sporns, O. (2005). Dialog: What are the potential principles for the autonomous development of value and motivation? *AMD Newsletter*, *2*(1).

Steels, L. (2004). The autotelic principle. In Fumiya, I., Pfeifer, R., Steels, L., Kunyoshi, K. (Eds.), *Embodied Artificial Intelligence, Lecture Notes in AI*, Vol. *3139* (pp. 231-242). Springer Verlag.

Thelen, E. (1995). Motor development: A new synthesis. *American Psychologist, 50*, 79-95.

Thelen, E., & Fisher, D. M. (1982). Newborn stepping: An explanation for a "disappearing reflex." *Developmental Psychology*, *18*, 760–775.

Triesch, J. & von der Malsburg, C. (2001). Democratic integration: Self-organized integration of adaptive cues. *Neural Computation*, *13*, 2049–2074.

Wellman, H. M., Cross, D., & Bartsch, K. (1986). Infant search and object permanence: A meta-analysis of the A-Not-B error. *Monographs of the Society for Research in Child Development*, *51* (3, Serial No. 214).

Weng, J., McClelland, J., Pentland, A., Sporns, O., Stockman, I., Sur, M., & Thelen, E. (2001). Autonomous mental development by robots and animals. *Science*, *291*, 599-600.

Weng, J. (2004). Developmental robotics: Theory and experiments. *International Journal of Humanoid Robotics*, *1*, 199–236.

*Tables:* See http://www.cprince.com/PubRes/EpiRob05/tables