An Early Investigation of Infants Learning

Jenny Cheung

A thesis submitted to
Auckland University of Technology
in partial fulfilment of the requirements for the degree of
Master of Computing and Information Sciences (MCIS)

2019

School of Engineering, Computer and Mathematical Sciences

Thesis Abstract

This research analyses and interprets psychological evidences of occlusion to investigate how an object could be represented and learnt that gives rise to infants' responses to the immediate environment from a computational point of view. Possible visual inputs are synthesized to construct a basic representation of an object that has not been addressed in the computational models reviewed. Followed by further studies of depth, mother recognition and cognitive development of other species, the basic representation construction connects with a straightforward process with desirable behaviours as output. Finally, tracking process, storage structure, representational input and process of responses are proposed. Future work includes development of an algorithm and implementation for verification.

Table of Contents

Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

____________________   _29 April 2019____________

Signature   Date

1 Introduction and Background

Artificial Intelligence (AI) has always presented a research dilemma for its researchers due to the ill-defined notion of intelligence. Consequently, AI researchers have often adopted a broad and loose definition of AI and this has led them to develop algorithms that are powerful but artificial. By artificial, it is meant that these algorithms could perform the task well but they themselves contribute little towards understanding of how the mind works. For example, many powerful machine learning algorithms have been developed (Graves, Wayne, & Danihelka, 2014; Liang et al., 2015; Shultz, 2012; Siegelmann, 2013) but the focus of researchers is to find ways that allow these algorithms to self-improve their performance on a specific task with each implementation. One popular approach is to train these algorithms with a lot of example data (Liang et al., 2015; Park, Kim, & Nagai, 2014; Petrosino & Parisi, 2015; Pierris, 2017; Pirri, 2011). Samples required range from hundred to thousand and took hours of iterations to "pre-train" the algorithm to its best fitness. It is obvious that infants don't learn by such a specialised training and they do not just improve their performance but gain an understanding of what they are doing. Consider another example, robot mapping. Robotics researchers have developed mapping algorithms that enable robots to learn instantly a precise map of a novel environment while exploring in it (Bailey & Durrant-Whyte, 2006; Durrant-Whyte & Bailey, 2006). However, animals (including humans) don't compute a precise map of their environments and are still able to find their way home using short-cuts. The question of how does the mind work becomes even more mysterious if we just look at a particular algorithm that simulates a specific feature of human intelligence.

To answer the question of how could we improve AI models, one must first locate the problem of current models. To find the problem of current models, the baseline of what is human intelligence has to be investigated. However, this becomes another mythical paradox. Jumping quickly to address the issues of some models and figure out a specific feature as solution could go even further away from how the mind works, if the works build on top of the ill-defined notion of intelligence. Instead of looking for characteristics or specific features of “intelligence”, this research would like to take a step back and carry out a critical review of existing works under the light of how does the learning process work in very early stages that supports the acquisition of these characteristics.

1.1 Thesis Structure

This thesis primarily identifies how (in behavioural terms) and what infants could represent from perceiving physical objects. In chapter 2, existing computational models of infant learning are reviewed to review the status quo of current findings and understand their limitations of representing physical objects. In chapter 3, the problems and research methodology are explained. In chapter 4 and 5, empirical studies of infant learning related to the problem of occluding objects are critically reviewed. In chapter 6, a brief sketch of a model of infant learning is presented. Chapter 7 will conclude the essential process of infants learning to build a computational model with a discussion of insights for future work.

2 Computational Model Review

This section reviews early computational models of infant learning with a focus of the inputs and the implication to the process. By reviewing these works, I would like to i) investigate the major concept of the models; ii) figure out what is inadequate to represent physical objects; and iii) analyse the aspects of empirical evidence required.

2.1 Constructivists' Schematic Models

Earlier computational models that followed Piaget's theory developed a learning model demonstrating a construction approach to build object knowledge and apply the knowledge to a novel event. Piaget (1954) suggested a mechanism driven by assimilation and accommodation. Infants absorb new knowledge and adapt to a novel situation using existing knowledge. Gradually, the knowledge accumulates with more sophisticated advancement of exploration. It is also found to be a generic theory amongst primates that bear similar transformation phenomenon (S. Wood, Moriarty, Gardner, & Gardner, 1980). Piaget's theory has been adopted by many researchers to construct learning mechanism in a progressive manner computationally.

The goal of these computational models simulates infant learnings by interacting with physical objects, predicting a result from previously "learned knowledges" and responding to a novel circumstance. A novel event may come with a different stimulus or a familiar object in a different situation. These models apply existing action to similar situation with novel objects to acquire a new combination of stimulation, action and result. The construction of this learning is structured as "schema". A different combination would give rise to a new knowledge in form of a context. A context holds objects and relevant descriptive information to represent various knowledge. As a result, these models argue to be able to learn adaptively and progressively with new knowledges of physical objects. However, they quickly leap forward to object manipulation mechanism with an action but ignored basic questions from the learning process – what could be an object?

2.1.1 Earlier Model of Virtual Baby

Model

Drescher (1989) attempted to create a virtual baby that can learn on its own with reference to Piaget's theory. The virtual baby has asserted Piaget's action driven schematic learning with stimulation, pre-defined action and result. The virtual baby has a bird's eye view to capture visual image. The vision field is divided into five foveal regions with black and white colors in 2-dimensional grid. It has a virtual hand to act against the object. Objects are uniformly sized and moved. This construction of object in the context is either being represented as ON and OFF without fragmentation of partial ON. These objects move to a grid in a discrete way such that a grid is either occupied or emptied and the same object do not present in two grids. Objects do not rotate. The arrangement of the agent and objects are analogous to chessboard movement moving from grid to grid. Snapshot is processed with edge detection, figure and background distinction. Then it is used to form object sketches, construct object appearance and position in the spatial world. A context consists of a list of objects and 16 visual properties for evaluation such as shape, texture, color, etc. This same structure is also constructed for the result after an action executed with the context.

Drescher (1989) simulated the learning process of assimilation and accommodation with a schema mechanism. The virtual baby processed knowledge with a schema composed of "context-action-result". This schema trio acts as a basic unit of the internal representation. If the conditions of a context are fulfilled, the schema activates and perform the action associated in order to achieve goal. Goal could be explicitly declared or intermediate one or delegated one from last activation. Actions are pre-defined. If new object appears, the virtual baby would explore possible actions to it. It would apply all possible actions onto it and discover the result to form new schema. Schema has initial properties as global setting. When it is activated by a stimulation, it would be instantiated with a set of local properties. Properties include statistics such as correlation, reliability, duration and costs. They are kept to relate an item and the schema. If an item is either recognized as ON/OFF with some more properties such as generality, accessibility and desirability. Since one schema might not be able to fulfill the end state such as a goal of placing an object from pt A to pt B, multiple schemas could be chained up as intermediates generated. Therefore, synthetic items are constructed by the mechanism itself for the chains to represent some intermediate states or conditions. This work provides a mechanism to process beyond primitive input from perception and create correlations synthetically before the final output delivered. Explicit goal and exploratory goal are both significantly driving schema activation. When there is conflict, this model constructs a synthetic element to create intermediate states to resolve. Differentiation could then be enabled by spinning off synthetic elements while generalization is deduced by frequency counts. Results demonstrate that this mechanism can respond to novelty situations.

Discussion

This model mimicked and implemented Piaget's schematic process. An object was pre-constructed and assumed to be constructible with the list of properties and a state of visibility. It is not clear how can such a visibility be derived with more object changes.

Drescher's model has segregated perceptual and cognitive processing. Perceptual process maintains and exploits "knowledge about objects and space" while schema mechanism does not know it so that cognitive processing could be maintained as generic process. This design nurtures an environment for cognitive process to be able to act generally across different sensory signals without a direct dependency on a particular type of inputs. Drescher's model described that the learning of an object stemmed from action-based consequences from primitive perceptual signals and invented synthetic elements to chain up schemas and resolve conflicts. The idea of synthetic element allows an internal representation that is not perceived from primitive input. Hence, it enabled more advance level of computational processing and created buffer to handle conflicts.

The model represented object as ON and OFF for its visibility in a context. But how can a context form with these values in the first place? The model has assumed that infants were born to see the list of properties the context hold and could represent the visibility of an object. How can visibility be determined? How could we bind an object with its visibility defined from direct perception? How could visibility be defined with object changes?

The schema mechanism is tied with motor action such as grasping and reaching. The model attempted to try all motor actions to explore objects. However, early infancy does not perform all actions to explore environments. They are being tied with limited motor capability. Infants are definitely not born with these actions at first as motor capability takes time to mature. If no motor action can be taken, nothing is being learnt from perception before they can perform some form of motor action. Is this an implication of something missing from the learning process before an action could be performed? What could be a response initially?

2.1.2 Dev E-R model

Model

Aguilar and Perez y Perez (2015) developed a dev E-R model that entails with constructive knowledge learning with "automatic generation of ideas" (Engagement) and "analytic evaluation and modification" of the ideas (Reflection).

The model constructs a 3D virtual world with typical real-world objects in grid. Objects move in pre-defined path and they are not occluding each other. The vision is installed with a camera taking picture in 180 x 120 pixels with three color values, red, green and blue. Each picture is analyzed with the size and color differentiation of different luminous spots. Size starts with discrete value as big and small initially and differentiates into smaller or bigger size progressively; similarly, color is decoded with red, green and blue values. Novelty item of both properties are not automatically "learnt" but must be stimulated until a pre-defined threshold is reached. These features of color and size would form the contents of a current context which describes current perception; it is used to search for previously developed schema. Current context would hold: i) features of color, size, movement and position of the object, ii) expectation, and iii) three additional drives of affectivity, emotions and motivations. The context alone could describe an object without a dependency on an action. Then, a basic schema could be formed with a context and an action. Dev E-R model may not necessarily tie with an action by the agent but could be any random action imposed. Once an action performed and the result is obtained, a developed schema would be constructed with context, action and expected context.

This model starts with how a new feature in a context could be recognized instead of automatically registration of a property. Certain threshold must be reached to learn the feature. This enables a context construction with "learnt" properties and can initiate learning of an object without an action compared with Drescher's model. Context could live without action and serves as a description for current situation. Therefore, an action is not mandatory to hold an episodic description. An object could initially be learnt by observation.

Similar to Drescher's model, development of knowledge is earned by simulating accommodation and assimilation processes with internal drives. Motivations drive the learning with goals. Affectivity shifts the attention by pleasure for luminous spots and displeasure for disappeared spots. Emotional preference of interest, surprise and boredom are hardcoded for object appearance and disappearance to guide an attention. Learning of knowledges starts when an object in interest is recovered by chance and the model would derive an action such as a gaze to the right for the recovery as an abstract schema. If the action also recovers other objects, it would be generalized in a schema. For an expectation is not met by carrying out an action, curiosity in "motivation" would cause the schema to modify and become more detailed and specific for more exploration. The model would try to land at equilibrium such that schema is stabilized with most of the successful prediction or expectation fulfillment. When a schema stabilized, knowledge learnt previously in form of schema would not have to be 100% matched with current context. The model will perform a random choice of either fully or partially matched searching of developed schema. If engagement of developed schema failed, it would automatically decrease the matching threshold such as another developed schema with dimmer spots, a smaller size or different level of affectivity. Selection would also be based on success rate of prediction if more than one schema matched. Then, if the expectation is satisfied, a new schema could be spin-off to differentiate.

Discussion

This model is another model developed from Piaget's schema. It has similar problem as Drescher's model that has not addressed the problem of what is being represented from immediate environment.

The model has not specified how could an object be constructed in a context. Current context of the model held a couple of features including movement. If an object moved and became invisible, how could the object be described? Does it even being represented in the same context? Regarding motion, similar to Drescher's model that this model also held movement as an attribute in a context. Is movement a feature to be described in a context? What could movement mean for perception? The model assumed that infants were born with movement classification mechanism. To describe an object, this model measure positions of an object with absolute value and aligned the absolute positions with relative size description. If size was described as relative measures, why would position be absolute?

This model does not actually clarify how an object could be fitted into the list of properties in a context though it enables feature registration. Without occlusion, object comes naturally as a complete "whole" thing and the model might have the same issue of not discovering a partially hidden object from perception. The changes of an object in different timeframe have not been considered as it is defined to be invariant perceptually. It is still obscure about what could be learnt from overlapping and dynamically changing objects. How can a full object be recovered from occlusion with this model? This might be assuming that an object, even as luminous spot, must be perceived and visible in full to refer to the same object. It seems to be very difficult to develop the model further to cope with visibility changes of physical objects.

2.1.3 Other Models on Structuring Schema Development

Models

Additionally, other research has supplemented further mechanism for both bottom-up and top-down schematic development in a tree structure. Other than differentiation between previous and current context, adjustment and integration could be added to the learning functions with bottom-up approaches processing perceptual signals and also top-down mechanism to apply knowledge to new perceptual signals.

Perotto and Alvares (2006) adapted a similar schema model for anticipatory learning. They differentiated perceptual details to generate new schema when previous knowledge is too broad. By this, a more specific schema is grown under a tree structure. Contrarily, if a schema previously learnt is too specific, adjustment would be made to existing schema and give rise to a more generic one grown over the top. With similar schemas, integration would be made. This provides a more complete set of functions in terms of internal manipulation of the knowledge development. Bi-directional growth would allow schema development to growth in flexible way giving rise to a network of knowledges.

Berthouze, Prince, Littman, Kozima, and Balkenius (2007) ride on a similar structure and go further to represent abstract unseen properties with the context-action-expectation vectors. The context vector represents a group of similar applicable situations, so as action vector. It has similar way to define the elements but allow an undefined state. As such, initial bootstrapping expectation, i.e. nothing is expected, or everything is a match, is possible. When the function failed, a synthetic schema would be created to deal with conflicting and inconsistent state with previous knowledge.

Discussion

All these further works have intended to address the unstructured development of schematic models that might cause problems of inefficiency. They could not completely resolve the mystery of object perception by the schematic development. Being efficient does not necessarily tackle the problem of representing an object and behaves as if perceiving an object. If an object moved, changed and partially hide, these changes do not even require an action to be performed. How could these changes be learnt observationally without motor actions? To be more efficient does not guarantee a resolution dealing with such a problem.

2.1.4 Constructivist Learning Architecture Schema Mechanism (CLASM)

Model

Following the principles of information processing approach, Chaput (2004), considering the work from L. B. Cohen, Chaput, and Cashon (2002), have constructed a Constructivist Learning Architecture (CLA) with Self-Organized Maps (SOM) that intended to implement the Information Processing Principles (IPP) of: i) knowledges emerged hierarchically from lower level to higher level of units; ii) no definite boundary in the hierarchy enabling a continuous growth of the levels; iii) learning mechanism inclines to start at highest level available or otherwise fall back to lower level to accommodate; iv) the learning strategies applies across various domains continuously. SOM is a hierarchical neural network that requires training. Training allows prototypes to be formed in a map. Each of the map could utilize nearest neighborhood choice of representation as well as frequency to locate a suitable cluster of features. Features are arranged in layers of SOM and designed with hierarchy such as from low to high level, specific to abstract. The SOMs are then combined with Drescher's schema mechanism following the principles of CLA to form CLASM. CLASM supplements the missing process in schematic growth such as pruning intermediates, harvesting and freezing matured schema, and claims to be more efficient.

CLASM works with the schematic mechanism to hold object properties as prototypes of various features and grow the knowledge in hierarchical manner. It starts with an action SOM holding a vector to store states of items (on/off) and result lists. Result items must be consistently reflected to be represented with a weighting in the final matured schema for harvest so that all positive items have to be on and all negative items have to be off. In case of unexpected result, novel result triggers a new schema creation and chains with other reliable schemas. Schema chains backwards to search for valuable items. For example, an object entered visual field stimulating a subsequent action without a centralisation; weak or poor sensory information could be further rectified to provide new information via synthetic elements to reach the goal. Intermediate schemas could be synthesized, like Drescher's work. The model has replicated and extended the functionality reported by Drescher's schematic mechanism.

CLASM implements IPP efficiently in a hierarchical manner. The mechanism of fallback to higher level (with more abstraction than precision) provides a good way to accommodate new changes that have not been discovered yet for an object but largely similar. In other words, the features from inputs are still referring to a similar boundary of "object", not an entirely new discovery. Pruning, harvesting and freezing mechanisms supplement supporting instruments as to the threshold accumulation of object features learning in dev E-R model but in knowledge level.

Discussion

CLASM could not represent physical objects in the process of infant learning better than previous models. The model assumed infants were born with accurate prediction capability or received an intensive training before they can learn. The vision fed to SOM, that was trained as a neural network, was exploited for accurate prediction of location of an object or object properties. It is doubtful that infants owe abundance of resources for accurate prediction initially but not mature in a progressive manner. Under such a constraint, they could still exhibit a responsive reaction towards various circumstances and demonstrate improved responses and learnings progressively, from movement changes to causal relationship. They don't watch hundreds of similar episodes before a response. It might be arguable that infants might have rich visual experience but the perceptual representation from their visual experience is different from specific training requirement. It seems to be difficult to work out a representational process for immediate environment in a general manner with SOM. To simulate different behaviors for various experiments that are being shown in other models, CLASM has different design with different training requirements for its neural networks. In other words, an untrained object cannot be not represented and developed further. Hence, a model trained for specific purposes cannot be claimed as a model of infant learning. This implies engineers have to pre-define various learning episodes first before a representation could be constructed. How many episodes of physical world can be defined? Is it even exhaustive?

The model also carried insufficient emphasis for visual processing with the temporal factor and ignored how can representation be constructed for a moving object. We are unable to have more insights against the role of perceptual system contributing towards visibility of an object and its changes during motion. Though CLASM could rectify poorly perceived snapshots, do infants rectify poorly perceived snapshots as well? If they cannot see clearly, on what basis do they rectify the perception if they do not know what it is? The model is assuming an inborn capability to rectify poor vision to a state that is assumed to be true. Can we make such an assumption?

With all these pitfalls, CLASM cannot fully explain how objects can be represented with very young infants. The model has skipped the problem of representing and processing primitive perception in early infancy in a general manner. From the perspective of what can infant perceive initially, CLASM cannot provide more insights.

2.2 Trajectory Prediction with Neural Networks

Model

Munakata, McClelland, Johnson, and Siegler (1997) exploited a recurrent network to predict the position of a ball with a moving occluder. Both objects move back and forth in the visual field. The network is firstly trained with a single ball unit with a static occluder. Then, it is applied to predict more complicated situation such as both the ball and the barrier are moving. The prediction is made possible by forming an internal representation of the pattern of the activity, i.e. the trajectory pattern, and therefore could be used to predict trajectory in a novel situation. Trained network would be sensitive to expected outcome and therefore could exhibit an insensitivity to surprises. This behaviour is being argued to simulate a longer looking time in empirical findings. After internal representation of the pattern is formed, the predictability and sensitivity that are gained from trained trajectory pattern would be used to predict the trajectory in a new situation.

This model has demonstrated a learning of internally represented trajectory and argues empirical findings might be explained with the gain of accurate prediction of motion. An accurate prediction could be reflected by the sensitivity of the network. The object is represented as a dot product. The model track the dot's appearance and disappearance and concluded that the behaviours of infants might be caused by tracking.

Mareschal, Plunkett, and Harris (1999) has also built a connectionist model to track object under occlusion event. It is designed with three major functional modules: i) object recognition; ii) trajectory prediction; iii) response. Object recognition has divided into spatiotemporal and featural processing as like the ventral and dorsal pathway design of human. Images are taken at an interval and received at 4 x 25 grids of retina and features of darkness, contrast, hotness and hardness would be detected. Then the object would be represented by the recognition module as a spatially invariant represented object using unsupervised learning algorithm. After the object being recognized, the spatial information would be extracted to predict its next retina position. Both modules would then be consumed by the response module to output a voluntary action.

This model completes the picture of learning from visual perception as compared to Munakata et al. (1997) that it has taken into account spatiotemporal completion and representation of an object. It has considered that a fully hidden object could be represented internally beyond direct perception. The model addresses what infants preferred to look at in its recognition module, for example, some interesting features instead of the whole snapshot. Detection of the next retina position is incorporated as a perceptual process; it engages the agent in the process relatively to the object in addition to the factual spatial configuration between objects.

Discussion

Trajectory prediction models have considered motion as an important element in visual learning that objects might appear to be disconnected under the condition of occlusion. They tackled the problem of tracking the object without direct perception and suggested a solution to represent object in motion. These models assumed: i) object is invariantly represented in a trajectory prediction; ii) object tends to be contiguous with itself. However, objects that are partially hidden can be segregated initially. The assumption of contiguous object properties could be challenged. An object might be invariant along representation but what happens if the properties changed during occlusion. Is it an error? By tracking down the group of "features" and their relative alignments and analogies, it might be representable internally and predictable for corresponding motion path. The biggest hiccup would be the assumption of invariant representation that limits the flexibility of the internal representation development in the case of animated objects. It is wrong to assume an object can always be perceived in full in any given moment. This issue must be addressed before building any learning algorithm that claims to learn the immediate environment.

2.3 General Discussion

Schematic Learning

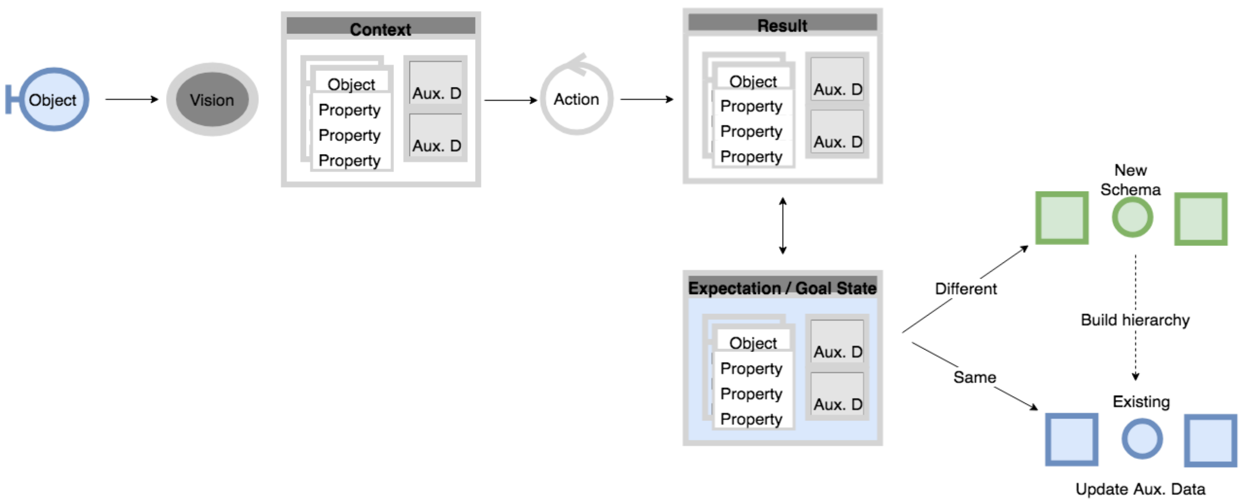

Figure 1 Constructivists' schematic mechanism

Schema models from constructivists focused to weave knowledges with prior learnings and use them to develop more knowledges with novel events as shown in **Error! Reference source not found.**. They provided a structure to describe perception in a schema (Aguilar & Perez y Perez, 2015; Drescher, 1989) and to learn its state changes with respect to actions structurally and hierarchically (Berthouze et al., 2007; Chaput, 2004; Perotto & Alvares, 2006). They are able to build more knowledge without imposing a boundary of what is adequately learnt and capable of expansion in both top-down or bottom-up approach.

Schematic models featured with trio construction of inputs-action-outputs very often tied up with motor actions. Action-based schemas have not addressed the problem of limited motor actions a newborn can react especially in first few months. Without those actions, objects can change on their own too. The changes have to be represented without an initiation of a motor action as well.

These models also held objects with a fix list of properties and may even enclose a state of visibility. Even if property has to be registered in the object holder as in the work of Dev E-R model (Aguilar & Perez y Perez, 2015), it is however not clear how these properties could be used by infants that objects could be changing and overlapping each other. Are properties received directly from primitive perceptual values? An invisible object that is being occluded with another object is not a recovery process of a poor vision. There is something seen, i.e. an occluder. The occluder is definitely the inputs for the snapshot but it is another object. Where is it?

These works have not shown infants learning in its early days about how an object could be represented with the decided attributes under different circumstances. It is also wrong to rectify poor perceptual signal automatically to avoid the problem that infants were born with poor vision initially. Any mechanism worked with the assumption of precise vision and mature motor capability cannot claim to be a model of visual learning of early infancy. The key importance is how could objects be represented for learning with poor vision if we are studying newborns. If the object cannot be represented as described in a context, how could subsequent schematic action be performed and give rise to learning? If we are not sure an object could be seen with these properties by infants, the schema mechanism may not work. Therefore, these works cannot explain how infants learn initially. Taking care of initial object representation is a fundamental issue to be resolved before the mechanism can be used to learn like an infant.

Learning from Motion

What is being learnt from a changing and moving objects in these models is ignored. Segmented features spread in a couple of snapshots considering temporal factor has not been incorporated into constructivists' models. Context holding objects in schematic models had taken object motion and object changes lightly. In fact, it may be due to the difficulty to recognise a non-static object that is not moving in discrete motion. Construction of object knowledge is too dependent on immediate environment available and very few of them could explain how object could be constructed from snapshots of varying primitive signals. For a context holding items directly, it is simply specifying a particular snapshot of an object to further develop. For a context holding properties describing a bunch of objects, it might build up different contexts from changes perceived and might not be converged to build up the same knowledge. From the perspective of infant learning, this is not an efficiency issue. This is a problem of ignoring the importance of construction of a representation for an object.

Guerin (2011) criticised that computational models have not been researched enough at analogous comparison and matching of previous knowledges. This could be due to a pitfall towards basic object understanding to construct a context and avoid building a process for dynamically changing objects such as human beings. Arbitrarily defined properties and their accommodating comparison mechanism might be a barrier to develop further. Therefore, the hiccups of the models created problem of matching issue. This can be a problem with models that assumed invariant object representation.

On the other hand, motion tracking models have point out that learning of dynamic objects initially might start with keeping an object within sight. This might imply a consideration of the subject for the purpose of focus calibration. When the position of the object changed, its motion has to be computed relative to the difference from previous snapshots. Subject could then respond with reference to a relativity measure of the difference. Temporal factor has to be considered to deal with snapshots and this has been found missing from some constructivists' models.

Even some of the connectionists models exhibited the ability of outputting a hidden object behind an occluder during motion (Franz & Triesch, 2010), but it could not be mistakenly treated as the object being learnt. It might reflect the fact that tracking of an object when it is hidden consumes a representation of an object as an input. This echoes constructivists approach of having an internal representation consumed for further knowledge growth like synthetic elements (Drescher, 1989). However, connectionists' models have not demonstrated the use of representing an object during occlusion. It seems to be a mean of showing its existence beyond direct perception only. If an object is hidden, what computation can be done with the representation?

The problem with trajectory prediction is that the motion is predicted for accuracy. Franz and Triesch (2010) worked out that a pre-training of random trajectories stimulation imposes 62 habituation trials to track accurately. The author pointed out that this number is still considered to be very large compared with empirical experiments. Does tracking an object aim at accuracy? Do infants track an object to accurately follow it only? What is the use of the representation when an object is hidden? Therefore, trajectory models aimed at accuracy cannot explain the learning process.

From all these models, they are unable to shed insights on the behaviors with dynamic and changing environment that an object can be sometimes invisible. For objects that are sometimes invisible and changeable, these models cannot explain what could be the necessary inputs to the process and why are they required. The assumption of having them as a fixed list of properties for all objects might remove many possibilities that are supported by the cognitive process.

Models that are action-base have not supported object changes to the inputs just by observation. How could a change induce a response of gaze shifting? What constitute such a response? It could be the changes of one or more objects revealing more of its properties and its correlation with immediate environment that give rise to knowing its propensities such as causal relation. These propensities might subsequently be associated with a cluster of properties and a group of other objects relationally, episodically or representationally. These substances might define the meaning of an object. How could a learning process make sense of perceptual inputs to respond? This is not just a top-down or bottom-up cycle of pruning or expanding objects in a context but a continuous exploitation of the relations between various properties with the agent.

A learning model should embed with a "mind process" that supports multi-facet growth of existing knowledge from object substances which are changeable, interactive and relative to the subject and its surrounding environment. The importance of maintaining an internal representation of an object might not only refer to rigid properties captured from direct inputs and might carry adaptive changes with respect to surrounding or self-initiated stimulations. If vision was poor initially, precision might not be the key important inputs for learning. Properties in an instant of an object could change. Change induces novel learning to the subject about the object in various circumstances. These learnings come together and could contribute towards more learning of an object without a defined boundary of what it is. Therefore, tracking and predicting the motion of an object is not just an output of accurate response, nor an expectation of an action, nor a mechanism built on top of a group of pre-defined properties; they could be the necessary mean to deal with various perceptual inputs. Through changes, what could be seen and represented is worthwhile to examine further.

3 Research Question and Methodology

3.1 The Problem

Understanding how the mind works forms the cornerstone for understanding how we become an intelligent species. As mentioned in introduction, the dilemma AI researchers faced is exactly the problem of limited understanding of intelligence itself. Mind process is mythical. Instead of just riding on ill-defined notion of intelligence to develop powerful algorithms, could AI research develop models that provide insights into how the mind works? Observe that nature's solutions to its problems are surprisingly perplexing and consequently, cognitive researchers often ended up discussing a paradoxical problem rather than identifying a clear solution to the problem. For example, how infants learn language is described as a paradox (Chomsky, 1979), namely Baker's paradox:

> "The discovery of the richness of the implicit knowledge of language immediately raised the question of acquisition. How can it be that every child succeeds in acquiring such a rich system so early in life, in an apparently unintentional manner, without the need of an explicit teaching? More importantly, the precise study of fragments of adult knowledge of language quickly underscored the existence of "poverty of stimulus" situations: the adult knowledge of language is largely underdetermined by the linguistic data normally available to the child, cognitive capacity must involve, in the first place, ..." (Chomsky, 1979, pp. p. 5-6).

It has been well observed that the input (i.e. the sentences that infants encountered) is not rich enough to support the grammar that infants use when they started speaking the language. Consequently, cognitive researchers debated whether grammar is learned or innate. Similar debates also occurred in spatial cognition. While empirical evidences suggested that we have "a map in our head", the nature of such a map remains elusive. For example, it is incomplete and imprecise and if so, how could it be used to orient oneself or to find short-cuts (Yeap & Hossain, 2019)? Despite the empirical evidence, cognitive researchers still debated whether a map is computed or not (Bennett, 1996; R. F. Wang & Spelke, 2002).

Developing a model for a paradoxical process is problematic. If AI researches were to contribute to our understanding of how the mind works, AI researchers need to develop models that help resolve these controversial issues raised in the cognitive sciences. To do so we need to analyse empirical studies of these cognitive processes to show what could be an appropriate initial process model for each of them. Yeap and Hossain (2019) showed that one could then empower a mobile robot with each of these basic processes and conduct experiments to show how various empirical findings about these processes could be reproduced in the robot itself. In this thesis, such study will be conducted that relates to the process of early infant learning.

But first, could infants learn? Early researchers believed that infants were born with "*blooming, buzzing confusion*" as their first experience of the world (James, 1890); it is stated that

> "The baby, assailed by eyes, ears, nose, skin, and entrails at once, feels it all as one great blooming, buzzing confusion; and to the very end of life, our location of all things in one space is due to the fact that the original extents or bignesses of all the sensations which came to our notice at once, coalesced together into one and the same space."

However, more recent research has shown that infants could respond and learn from interacting with their environments. Species and things could thus be linked in a complicated network; subject and objects are being differentiated, captured and reflected to the process consciously (Piaget & Cook, 1952, p. 19). Some researchers refer this as a capability of being able to infer and predict the properties and next states of immediate surroundings (Fitch, 2014; Spelke, 1994). Others refer this capability as i) being able to communicate and engage; ii) supporting embodied system such as organs, memory and learning; iii) progressing adaptive level-up from previous interactions with environment (Zlatev, 2001, p. 161). These suggestions point at a comprehension of an agent's immediate environments. If this is true, what do infants perceive of their immediate environment? How could infants comprehend their perception? What would be a basic computational model of early infant learning?

There has been much research now that demonstrates infants could learn much about the world they are born into within the first few months. These knowledge includes ideas about object permanence (Baillargeon, Spelke, & Wasserman, 1985; Bremner, Slater, & Johnson, 2015), occlusion (S. J. Hespos & Baillargeon, 2001a; S. J Hespos & Baillargeon, Cognition/2008; Wilcox & Baillargeon, 1998a), obstacle (L. Kotovsky; Laura Kotovsky & Baillargeon, 1994, 1998), support (S. H. Wang, Zhang, & Baillargeon, 2016) and others. In this thesis, the focus will be developing a process model for what could be learned initially about objects in view. How does the notion of "object" form in the infant's mind? What key processes are available at birth? To find answers to these questions, psychological experiments will be reviewed on infant learning and identify the important computational questions that need to be asked when developing a basic process model of infant learning.

3.2 Methodology

Traditionally, researchers define a broad question that lead to a study of relevant literatures. Subsequently, through discovery of existing knowledge of the subject of interest, researchers determine the gap for the research. However, be it descriptive, rational or a causal question, finding a gap from existing literature does not necessarily generate interesting research questions. Researchers faced paradoxical dilemma of being too subjective to describe a situation in a review (Kangasniemi et al., 2013) or asking less impactful questions (Alvesson & Sandberg, 2013). Alvesson and Sandberg (2013) suggested that the essence of asking interesting research question is to challenge existing assumptions and look for problems in those assumptions; by this, one could deviate from endorsing what has been known and contemplate a problem in a different way to yield an impactful study.

The research method adopted in this work is one based on a critical analysis of psychological experiments on infant learning. In my review, I pose a set of questions different from those of the psychologists. I am not concerned with the validity of their research or their theorizing of their work. Instead, I focus on the behaviour observed and a computational model that could explain that behaviour. Psychologists tend to provide a theory of infant learning in terms of what concepts are learned. As noted above, they argue that infants could learn about object permanence, occlusion, obstacle, support, and others. However, given that these are semantically laden terms, it is not clear what exactly is learned. In contrast, the approach here is to provide an explanation of what is learned in computational terms.

Marr, an early pioneer in AI research, developed a computational approach for studying biological processes. Marr, in his work of vision research, suggested that:

> “vision is primarily a complex information processing task, with the goal of capturing and representing the various aspects of the world that are of use to us. It is a feature of such tasks, arising from the fact that the information processed in a machine is only loosely constrained by the physical properties of the machine, that they must be understood at different, though interrelated, levels” (Poggio, 1981).

Marr suggested three levels of resolving the problem: i) computational theory, ii) algorithm, iii) implementation. The work here is done at what Marr refers to as the computational level.

3.3 Scope and Boundary

As noted above, psychologists have argued that infants could learn various concepts such as object permanence, occlusion, obstacle, support, and others in the first few months of their life. In this study, I will focus on how infants learn about occluding objects i.e. those that are not constantly in view. The bulk of the work presented here is thus spent on analysing related psychological experiments. I began with an analysis of a set of experiments directly related to the occlusion problem (chapter 4) and then I analyse further a set of experiments that are related to this problem in some interesting ways (chapter 5). For latter, I consider how infants and other species recognise their mothers and how infants see depth. Due to time limitation, only a brief sketch of the process model of infant learning is then presented (chapter 6).

The literature research was performed using three sources: i) library subscriptions of Auckland University and Technology; ii) Scopus; iii) Google Scholar. Keyword used for physical object learning include “infant learning”, “physical objects”, “occlusion” and “experiment or empirical”. For mother recognition, keyword “mother recognition” is used instead of “physical objects”. Literatures and their citations with empirical experiments of physical objects with infants would be included. Works that only interpret others’ experiments are excluded.

4 What could infant learn?

In this chapter, I review some recent empirical evidence supporting infants learning and in particular I focus on asking what computational questions could arise from these studies that could lead us to the development of an initial computational model of infant learning. In section 3.1, I discussed three assumptions I made concerning the infant learning process. In section 3.2, I reviewed the empirical evidence that supports infant learning. I focus on the notion of occlusion and show that what infants learn in these experiments need not be a semantic concept but rather a straightforward representation derivable from the input.

4.1 Three assumptions about what infants could see initially

If infants weren't born into this world feeling confused, what do they see initially? What basic capabilities do they have that would eventually allow them to gain an understanding of the orderly world that they are in. Based on the empirical studies reviewed below, I make three assumptions in developing a computational model of infant learning, namely: (i) infants could detect edge changes, highly contrasted features, shapes and other low-level features, (ii) infants do *not* perceive objects; and (iii) infants could track moving "objects".

Spelke (1990) describes the principles of object perception that i) objects segregate from each other with heterogeneous pattern visually; ii) objects are situated in "*rich and changing environment*" and beyond direct visible configuration; iii) visual perception is divided into arrays of information in a general segmentation process. These form a basis for visible direct perceptual inputs. If infants were born with confusion, what could they perceived and comprehend from such a complicated inputs to form something meaningful to them? The very first process is to make sense out of direct perception. Understanding direct perception could never be defined clearly when immediate environment is composed of many complicated surfaces composed of contrastable properties. What do infants understand from these properties?

4.1.1 Perceiving changes

What do infants look at when they were exposed to the immediate environment? The world is full of objects, both static and animated one. The first moment, their visions capture a snapshot of the environment, it could have changed in the next moment. Could they have learned in between or "in the transitions of these moments"? What do they initially attend to?

Infants respond to a change of continuation of an edge. An experiment conducted by Salapatek and Kessen (1966) projected a solid black triangle against white background to infants and captured their eye movements scanning the shapes. The gazing trajectory has been recorded. Results showed that all infants show a preference of scanning one of the vertexes of triangles instead of looking evenly across the lines of the shape presented. What makes a difference of vertexes from lines? It is an angular point joint by two lines, a change of continuation of an edge as opposed to a straight line. Infants seem to respond to a change in continuation of an edge.

Bronson (1990) has investigated 2 to 14 weeks of infants and showed a result very similar to Salapatek and Kessen (1966). Very early infants do not scan all the properties but the highly contrasted feature to fix their gaze. Even when they grow older and become more attentive to the contour or other details of the object, they would still switch to fix their gaze at the highly contrasted feature when the object is flickering. Therefore, infants basically utilise the strategy of fixing their gaze at the highly contrasted feature as a reference point to follow the motion.

Further findings found that infants respond to a difference in shapes more than contour density. Slater, Morison, and Rose (1983) tested whether infants' looking preference were a result of contour density or due to the novelty of the shape. New-borns were firstly habituated to one of the four shapes which are square, cross, circle and triangle, then they would be presented again with a familiar shape and another novel shape. Results show that infants look longer at novel shapes. Subsequently, the infants were tested again but with thin or thick contour of the black and white shapes that is slightly shifted in various angles. A similar result is obtained; infants were not responding to the differences in contour density but to a different shape.

Psychologists also reported that infants were found to shift from being sensitive to orientation change to an angle change between 6 weeks and 14 weeks years old. Leslie B. Cohen and Younger (1984) habituated infants to an angle formed by two lines first and then tested with stimuli of either same angle different orientation or same orientation but different angle. Results show that younger infants were not showing looking preference towards angle change while older infants were. The difference between orientation and angular test is that angular test has the shape with one of the line in common with habituation one. It seems to show that very early infancy respond to a higher degree of changes in terms of overlapped lines and surfaces.

The above experiments show that infants could attend to changes in geometric features. Computationally, one could argue that further understanding of their world could be built upon these initial exposures perceived but how? Responding to a change requires a comparison of two snapshots. What constitute a difference in this comparison by infants?

4.1.2 Perceiving objects

The immediate environment encapsulates a lot of potential changes, be it the lines and shades changes in a static object, or, a position and properties changes in a moving object. Odd shapes or surfaces are all possible. This section analyses those studies that investigate how infants perceive these changes.

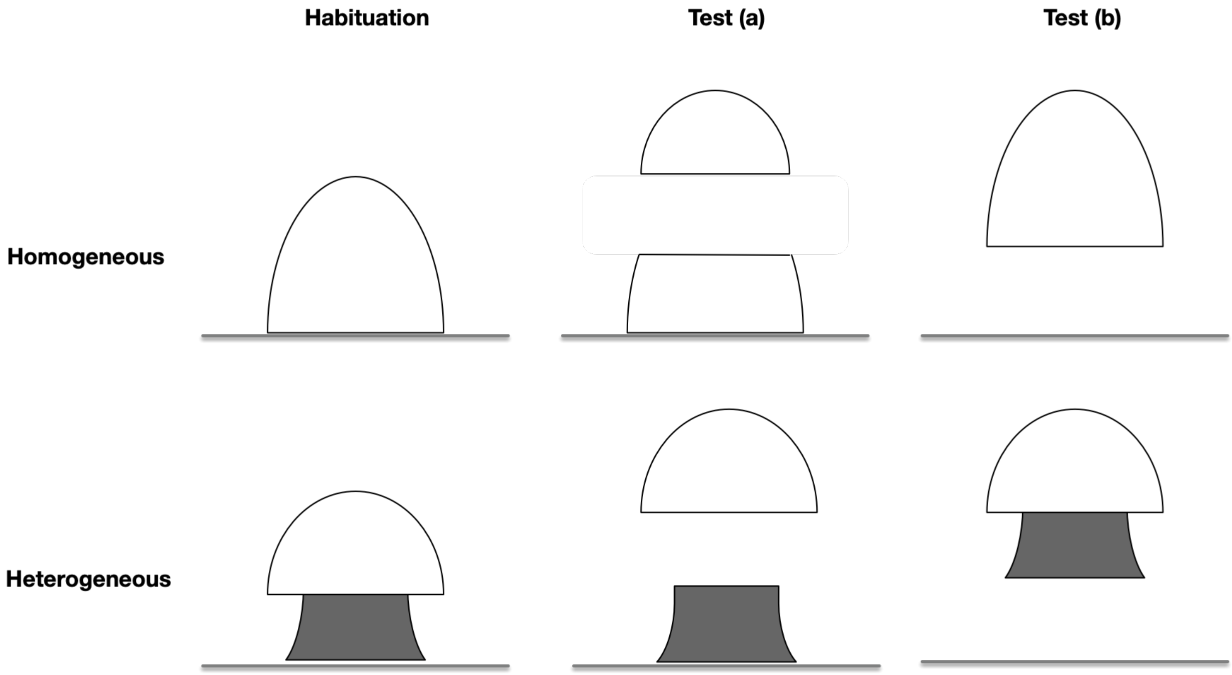

Figure 2 Homogeneous (top row) and heterogeneous (bottom row) display, both displays were tested with a lift on top. The object was either broken into two parts in test (a) or remained as one object in test(b)

Infants not just look at changes of lines and also shades and texture from a static image in early stage. Spelke, Breinlinger, Jacobson, and Phillips (1993) have examined the effect of colour and texture, continuation of features, good form (as opposed to odd shapes) for 3, 5, and 9 months infants. This research compared the performance of infants with adults. The experiment involved objects display that are either homogeneous or heterogeneous objects as in Figure 2. When the object was being grasped, it was either lifted into the air with the whole object like test (b) or broken into two parts like test (a) along the change line of the colour. Participants were firstly habituated to the object sit on the floor statically or just an empty floor only. Then, they were measured against whether the object is a one-piece object or two-piece parts by asking adult participants and measuring looking time for an infant. It is found that adult perceives heterogeneous object as two-piece parts while homogeneous one as one object. In contrast, 3-month-olds responded indifferently to both homogeneous and heterogeneous objects. 5-month-olds and 9-month-olds are however influenced by the effect of colour and texture, continuation of features, good form; they looked longer at the homogeneous object that separated into two along the discontinued contour and colour by the grasp. However, they didn't show a preferential looking for heterogeneous object that move as one piece.

This findings shows that 3-month-olds has not formed a different response to homogeneousness of properties. They have not learned that homogeneous object usually moved together. 5-month-olds do demonstrate that similar properties adjacent to each other usually group together in different snapshots, disregard of other changes. This caused a formation of a response to follow homogeneous properties as a whole. If the homogeneous object is separated into two, it was very different from most of their previous encounters of repetitive snapshots from 5-month-olds onwards. This implies that stationary surfaces with similar or contrasting shades or patterns cannot automatically earn them a definition of "an object".

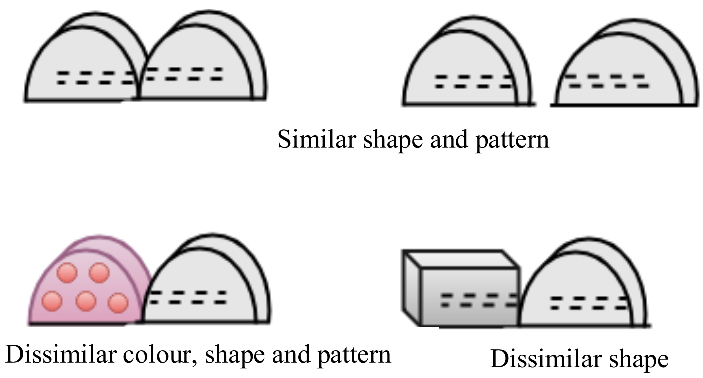

Figure 3 Pulling objects in different pattern, shape and colour

Some researchers investigated what could be the response with an integration of various visual information. Amy Needham (1999) studied that the effect of shape, colour and pattern amongst 4-month-olds infants as in Figure 3. Experiments established are similar to the previous one that habituated infants with stationary display and then pulled two objects with various shape, colour and pattern. Results show that infants look longer if the objects were separated by a pull with similar shape and pattern, moving together with dissimilar-shape and separation of dissimilar-colour-and-pattern. It seems that 4-month-old has not learned to connect to a significant comparison with colour and pattern.

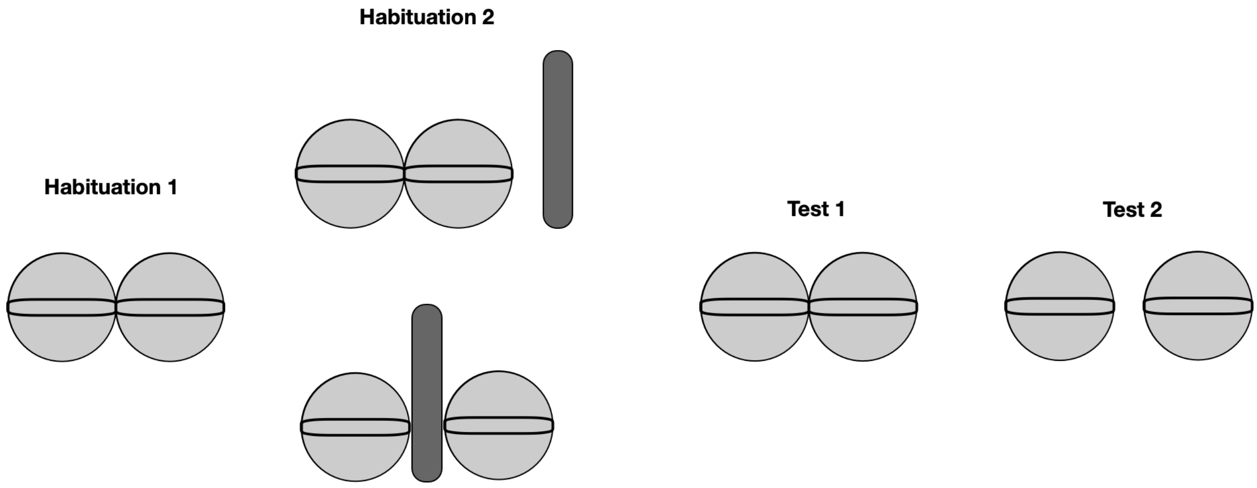

Figure 4 Habituation: firstly habituated with a static set (Habituation 1) and then habituated to the same set with a blade set aside or in-between (Habituation 2). In tests, blade was removed and the object was pulled at one side making it either move together (test 1) or broke apart (test 2)..

Amy Needham (1997) established an experiment on 8-month-old infants that initially exposed with two objects stick together and a thin blade in the middle or at a side of both objects. In a move-apart event, one of the objects was pulled apart while the other one remained static; in the move-together event, both objects move together as if a single object. Infants were habituated to the two objects stayed stationary first, then the blade would be placed accordingly for habituation as illustrated in Figure 4. Results show that infants who saw the blade being placed in the middle look longer at the object moving together. Those who saw the blade placed a side look longer at the object moving apart. This experiment reassure that surfaces with less change (an alignment) of lines and surfaces would be responded as if they belong to one under motion in older infants.

These experiments seem to suggest that even very young infants would be able to synthesize visual signals into connected group of properties and compare them with adjacent properties for similarity (A. Needham, Baillargeon, & Kaufman, 1997). Similarity of properties does not give rise to the concept of an object.

4.1.3 Tracking abilities

Gronqvist, Gredeback, and Hofsten (2006) found that infants have mechanism to handle horizontal and vertical component of motion and tracked the motion of an object with faster velocities than the velocities of the object. This means that visual focus (gaze) could be shifted to the next position even faster than the object moves to. From the perspective of physics, velocity is a vector that measures direction and the magnitude of position change taking a frame of reference against time. A frame of reference is an abstract concept describing one object's motion relatively to another object as physical science view everything as non-static. Infants' tracking behaviour is actually a directional change of the position of the object.

This implies that tracking must engage a coordinate system to infer next position by recent positional change. Tracking is composed of these two components and can be represented by a coordinate system to cater motion as a result of X-axis changes (horizontal component of motion) and Y-axis changes (vertical component of motion). The forward gazing mechanism indicates that gazing response could not be reactive by current perception but predictive one. This could only be possible if a response of next position is predicted reacted ahead of the movement of the object, i.e. move to the next position even before the object reached there. By that, tracking could move faster and keep following the motion path.

A comprehension of immediate environment involves diagnosis of a snapshot with varieties of surfaces and properties. As shown in the previous section, these properties do not automatically form an object even if they are visible. Making it worse, objects are sometimes invisible. Computing the contrast between different properties and forming surfaces cannot help to resolve the construction of an object that could be moving and invisible. To resolve the question of how infant learn to form a concept of "an object", occlusion events have to be studied to investigate what is seen and how to handle unseen.

4.2 Occlusion Events

Experiment 1 - A partly hidden object

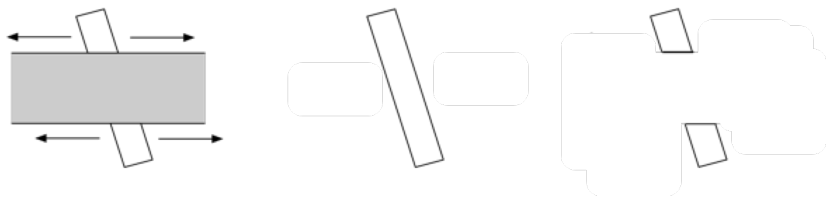

Figure 5 Adopted figure from Kellman and Spelke (1983) for rod and occluder test

In one of the early studies, Kellman and Spelke (1983), investigated how 4-month-old infants perceive the unity of partly hidden objects. The key idea in their design of their experiments is to have an object hidden behind an occluder and manipulated in various ways. The test is to see how infants react when the occluder is removed, revealing either (a) a complete object or (b) two disconnected objects (Figure 5). Their experiments are summarised in table 1. Infants were habituated to the moving, partly hidden rod and then tested with moving complete and broken rods. Result: infants perceive the rod as complete.

	Habituation Setting			**Result**
Experiment	**Rod/Triangle**	**Occluder**	**Position**	**Rod (Perceived as)**
1	Complete, moving rod	Stationary	Rod behind occluder	Complete rod
2	Complete, stationary rod	Stationary	Rod behind occluder	Indifferent response to complete / broken rod

	Stationary, broken rod	Stationary	Rod in front of occluder	Broken rod
	Complete, stationary rod	Stationary	Rod in front of occluder	Complete rod
3	Complete, stationary rod	Stationary	Rod behind occluder	Complete rod
	Visible, broken rod	Stationary	Rod behind occlude but the rod can be perceived as broken	Broken rod
4	Complete, stationary, triangle	Stationary	Triangle behind occluder	Indifferent response
	Complete, stationary, triangle	Stationary	Triangle in front of occluder	Complete triangle
	Broken, stationary, triangle	Stationary	Triangle in front of occluder	Broken triangle
5	Complete, moving rod	Stationary	Rod behind occluder	Complete rod
	Complete, moving rod	Moving together with the rod	Rod behind occluder	Indifferent response
	Complete, stationary rod	Moving	Rod behind occluder	Indifferent response

	Complete, stationary rod	Stationary	Rod behind occluder	Indifferent response

Table 1 Occlusion experiment by Kellman and Spelke (1983)

5 additional experiments were conducted to show: i) when the object behind the occluder is not moved, it is not perceived as a complete object on its own; ii) When the object behind the occluder is moving laterally, the object is perceived as a complete object.

Johnson and Nanez (1995) also performed the rod-and-box occlusion experiment in two-dimensional computer display and compared it with Kellman and Spelke (1983). Results of responding to a partially hidden object as full object with 4-month-olds were similar to Kellman and Spelke (1983) while 2-month-olds have indifferent response towards complete and broken rods.

Discussion

Properties (i.e. object that is not fully in view) move with the same magnitude, they are tracked as one object even if they are not fully visible. An object that is moving can be sometimes partially invisible. No perceptual signals are available for occluded portion. What could be the input for occluded portion?

At 2-month-olds, the moving object is not being represented for computation. Only the visible perceptual signals are being compared. The tracking path is being calculated from direct perception and projects a coordinate to gaze at with the occluder perceived. Nothing is expected for the hidden portion of the moving rod. When a complete or broken rod is being displayed, the occluded rod is just being updated with whatever being presented.

At 4 months old, an expectation of a connection between properties moving together could be formed. However, since unseen portion has never been seen, the occluded properties remains in question what it is. The actual visible perceptual signals are referring to the occluder and an internal representation with outstanding hidden portion to resolve. It is a combination of inputs from internal representation for tracking and relate its relative coordinates with another representational inputs of the occluder. When a complete or broken rod is being shown, the similarity of the properties relates back to the occluded rod and induce a "fill in the blank" action to answer the "questionable" invisible portion of the representation. Evidences suggest that the "fill in the blank" action tend to fill in the actual perceivable complete rod (i.e. properties are connected together) to the representation and becomes "one object" if no surprise was given. Why is a complete rod preferred?

Experiment 2 – Occlusion of various shapes

Psychologists have conducted many experiments to show that infants could learn much about occlusion events from visual perception. This section reviews such experiments and discusses what infants could have learned from an information processing standpoint.

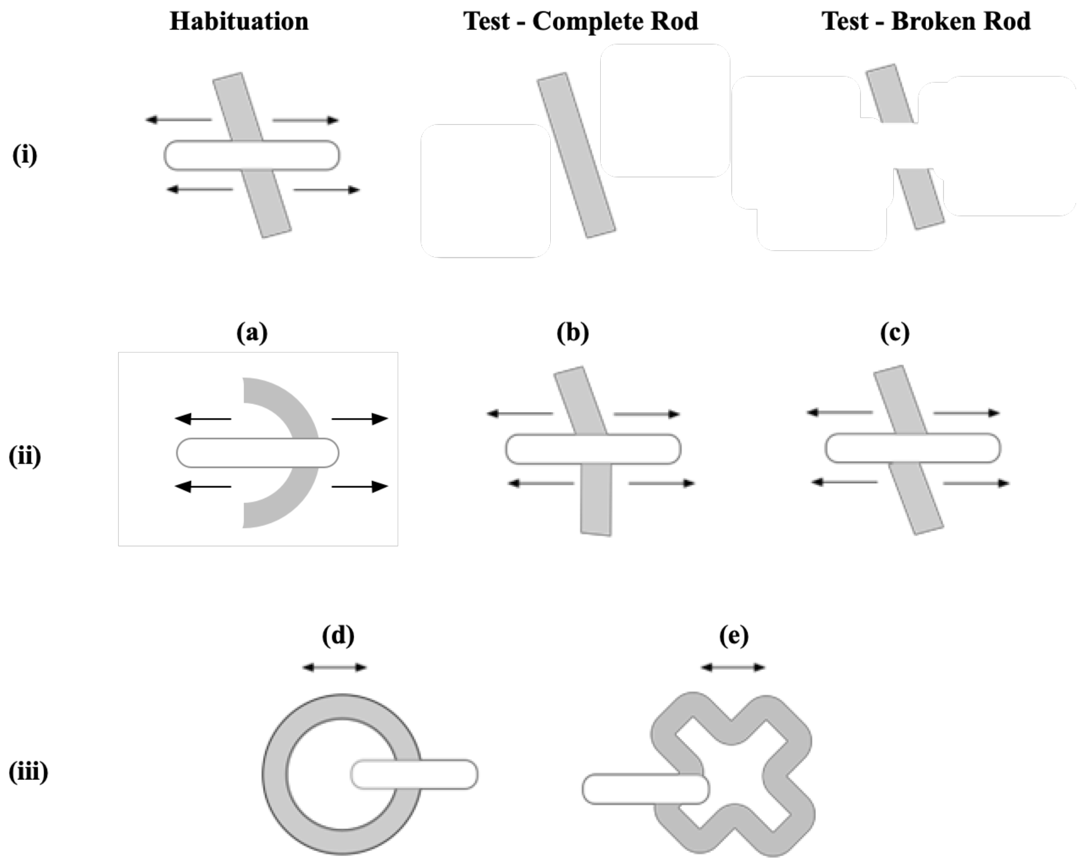

Figure 6 figure adopted from Johnson, Bremner, Slater and Mason (2000) and modified. Habituation was done with a laterally moving object with an occluder; then in test phase, looking time measured with complete rod and broken rod(i). (i) was repeated for all objects from (a) to (e) and tested if the occluded objects are perceived as complete shape or broken shape.

Johnson, Bremner, Slater, and Mason (2000) conducted experiments similar to Kellman and Spelke (1983) that habituate 4-month-old infants with a moving rod behind an occlude and then tested them with a complete rod and a broken rod 6(i) as shown in Figure 6. But they repeat the experiment replacing the habituation rod with 5 shapes as shown from 6(a) and 6(e) as in Figure 6. For 6(ii), it is found that infants look longest at bended rod 6(c), shorter at half ring 6(a) and immediately at bended rod 6(b). For more complex shape in 6(iii) such as ring 6(d) and cross 6(e), they looked longer at the broken shape as well. They responded to the occluded shape as if it was a complete object. This supports that an occluded object in motion would be viewed as one object.

Other factors such as texture, alignment and relatability of the edge also affect the comparison in a motion. Johnson and Aslin (1996) suggest further with occlusion tests that 4-month-old infants are capable to look at the object as if visible parts are connected, the rod parts could be perceptually relatable or they aligned under lateral translation. They studied against three conditions with the broken rod test ((i) in Figure 5) as in Kellman and Spelke (1983): background with a different texture, alignment and edge relatability of the object (relatability refers to that the edges of the rod intersect behind the occluder) of the edge of the rod. In the test of textured background, infants were habituated to textured or non-textured background with alignment and relatability being held as constant. It is found that infants who were habituated to textured background dis-habituated more to a broken rod (looking longer); infants who were habituated to non-textured background has indifferent looking behaviour against broken or complete rod. The research illustrated that a high contrast of the texture of an occluder from the background is required to respond to the occluded object as a complete one. This seems to echo that they attend to properties with highly contrasted feature. The contrast constitutes a difference of noticing the group of properties separately and makes it available to the comparison separately as another group of properties.

Discussion

At 4 months old, irrespective of shapes, properties moved together are being tracked as one group and are expected to be connected. All shapes involved in the experiment are perceived as complete shape. A complex shape requires a longer encoding time when a complete object appear. If the visible portion can be aligned with straight line, it took least looking time. This suggested that the invisible portion is preferred with less contrasting changes, i.e. aligned lines, same shades, etc., to update and complete the invisible part of the object.

Experiment 3 – A fully occluded object

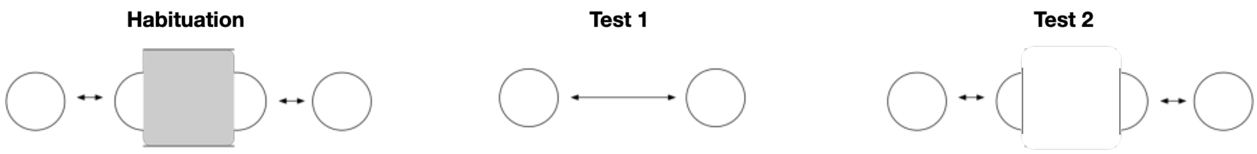

Figure 7 figure adopted from Johnson et al. (2003) for ball and occluder test

Empirical evidences suggest that a wider occluder could induce a longer looking time. Johnson et al. (2003) adopted a similar version with a rod as in Figure 5 and also run the test with a ball as in Figure 7. This research adjusted the width of the occluder. Infants were firstly presented with the occlusion event in habituation phase to familiarise with the display, then they were shown with a complete rod or ball event (test 1); that was subsequently followed with a broken rod or disconnected ball event (test 2). 4-month-old infants respond to the trajectory of a complete rod and complete ball event same as occlusion event (i.e. looking longer at the broken event) if the occlusion gap is narrow enough while 6-month-olds could do the same with wider occlusion gap. Could a faster speed remedy the situation?

Bremner et al. (2005) show that reducing the time gap by rolling the ball faster can facilitate the motion path to be tracked as continuous path. The experiment increased the ball size to about the same width of the occluder in the middle such that the ball entered the occlusion would quickly be seen again. This setting allows large spatial occlusion but small temporal gap. Result proposed that 4-month-old infants can respond to the object same as a complete object. However, in the other experiments in the same paper, more attempts tried to find out the relationship between the temporal gap and trajectory continuity, it does not have a significant experimental effect for reducing or increasing the speed of the ball. This inconsistency of the findings could be due to that up to a certain wideness of the gap, a change in speed does not constitute a major difference to respond with.

Discussion

The experiment indicates that the younger the infants, the higher tendency to skip tracking the object when it is invisible for wider occluder. The object could be tracked as one continuous path only if occluder is not too wide. When an object is moving in higher speed, the object can also be tracked as continuous path when it is invisible. However, speed has no effect to keep tracking the motion if the width of the occluder is very wide. This means the middle path is ignored and is directly jumped to the exit of reappearance of the moving object. For such case, no representation of the moving object is required as it does not matter.

Experiment 4 – Stroboscopic motions

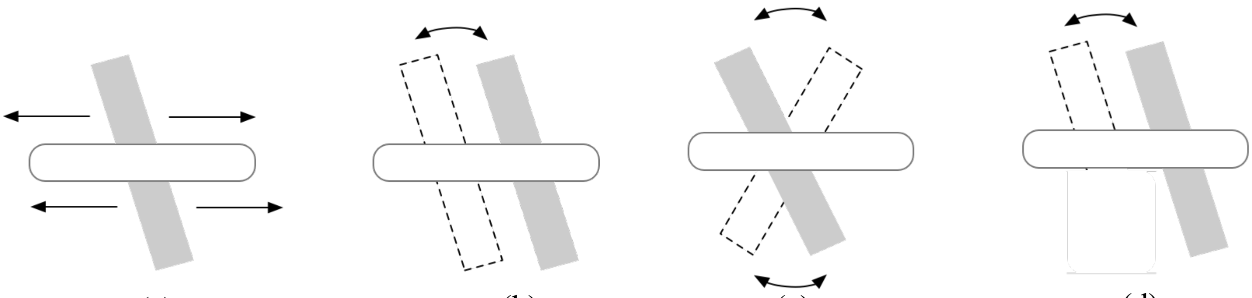

Figure 8 Habituation phase with various stroboscopic motions

Figure 9 Test phase with complete and broken rod display

Psychologists found that infants respond to discontinued snapshots of a moving object as if it is moving continuously. For occlusion event, new-borns were tested further under a stroboscopic motion and showed a response similar to a smooth motion (Valenza & Bulf, 2011; Valenza, Leo, Gava, & Simion, 2006); the research conducted with rods that were flipping with stroboscopic motions in various way as in Figure 8 and Figure 9. They attempt to test if a few days old infants respond to the snapshots of rods as if they were smoothly moving and swifting. During test phase, infants were demonstrating a different response towards broken rods as if habituation is the same as a complete rod. These experiments suggested that stroboscopic motions were perceived as continuous motions.

Discussion

Stroboscopic motion can be represented as smooth motion. This implies that the position between each snapshot can have observable discrepancy. The observable discrepancy does constitute towards a change. In (d) of Figure 8, the half rod and a complete rod are two different objects. Therefore, the half rod can neither being seen as a complete or a broken rod as both of them have a lower part. The perception of a complete rod refers to the tendency to connect the groups of properties that changed together.

Experiment 5 – Comparison with occluder

Baillargeon and DeVos (1991) tested 3 months old and 3.5 months old infants with a carrot and an occluder with an opening on the upper part. They suggested that 3.5 months old infants expect a tall carrot to be seen behind the occluder with windows but 3.0 months old infants did not expect the tall carrot to be seen even if the carrot is taller than the window. This lead to further testing of occlusion to verify if height of an object can be compared against the occlusion.

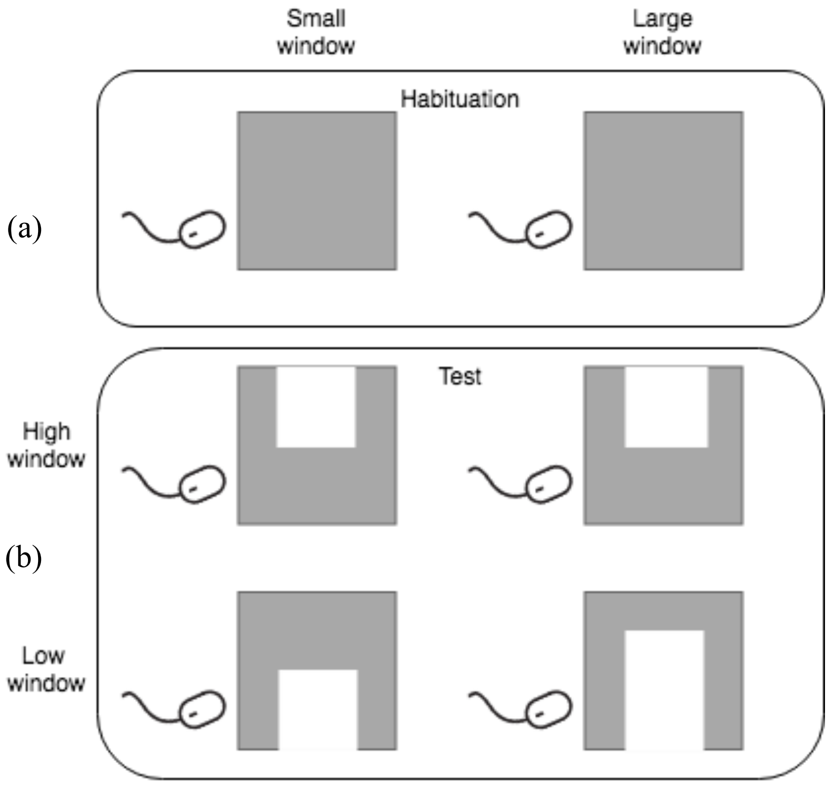

Figure 10 Figures adopted from Aguiar and Baillargeon (1999) and modified.
An object passing through an occluder. (a) Infants were firstly habituated with full occlusion and then (b) tested with an occluder with its middle section cut in 4 different ways.

Aguiar and Baillargeon (1999) established an experiment as in Figure 10. Habituation is executed with an object moving laterally behind a screen that is fully occluding (a). Then infants were tested against occluders with windows that was opened on either upper (high window) or lower part (low window). In low window events, the object did not show up when passing through the window. Results show that 2.5 months old infants did not have a preferential looking toward both high and low windows events. It seems that 2.5-month-olds did not look longer at the test event of low window.

Figure 11 Figures adopted from Aguiar and Baillargeon (1999) and modified.
Fully cut and disconnected screens test

A second experiment went on further to replace low window occluder with a fully cut screen in the middle and disconnected from both sides as in Figure 11. Infants were looking longer in the two-screen event than the high screen event with the object invisibly passing between the disconnected screens. The authors claim that infants were expecting the object to show up between the visible gap of the screens while remains occluded between one connected screen without other dimensional changes such as height.

Discussion

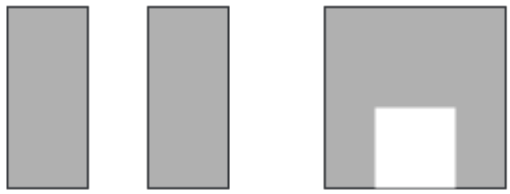

Figure 12 Two screens scenario (left), One screen with a window (right)

A moving object that is fully occluded can be processed as seen or not seen in a relation with the occluder. For 2.5 months to 3 months old infants, they can only tolerate a narrow occluder to represent the objects and compare with the visible inputs (of the occluder), otherwise, they will skip to the exit of the reappearance of the object directly (Experiment 3). Therefore, with the habituation of a wide occluder, they are already habituated to skip the tracking. During testing, the object is also being skipped for tracking and the window of the occluder is ignored. Older infants (after 3.5 months old) have a higher tolerance of the width of occluder, therefore they track the motion when the object is hidden and can use the representation of the object to compare with the visible occluder. In a two screen events, the width of the occluder is narrower by dividing the wide occluder into two surfaces (Figure 12). Therefore, this enabled younger infants to keep tracking the moving object and a representation of the object was constructed for motion tracking when the object was hidden.

Experiment 6 – Validation of position

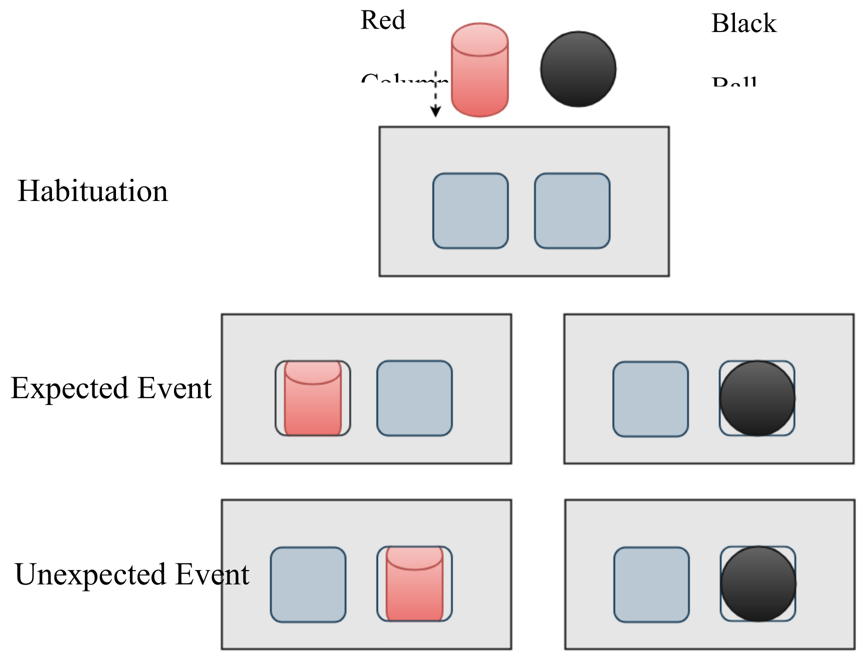

Figure 13 figure adopted from Wu, Luo, and Ballargeon (2006) with modification. Habituation: a red column and black ball are lowered into the stage with two windows. Tests of expected event: the windows of the stage were opened showing a red column on the left window and a black ball on the right window. Tests of unexpected event: the windows of the stage opened twice, first time shown with a red column, closed, then opened again to show a black ball..

Wu, Luo, and Baillargeon (2006) examined that 4-month-old infants respond to a different object if its position was not consistent with the movement direction. The experiment was conducted with a fully covered stage opened with two identical windows that could be lifted from the top of the stage to reveal the content behind as shown in Figure 13. Infants were habituated to the situation that a red column and a black ball are lowered from the middle of the stage into the covered area. In an expected event test, right and left windows were opened to reveal the column and ball respectively. In an unexpected event test, right window opened twice with first time revealing a column and second time a ball. Infants looked longer at the unexpected event. They suggested that objects occupied different space could be understood by infants.

Discussion

The last seen coordinate changes can be used to compute the position of the hidden object and compare it with what is being revealed. This experiment shows that even infants can't see the red column, the red column can still be represented with the possible coordinates. There are two possible representations available during occlusion, positional changes and properties of an object. In this experiment, the positional changes (during visible period) can be exploited to compare the appearance of the object. When the red column reappeared, the perceptual inputs are validated against the properties and positional change and therefore formed a surprise when it is inconsistent.

Experiment 7 – Differentiation of occluded object

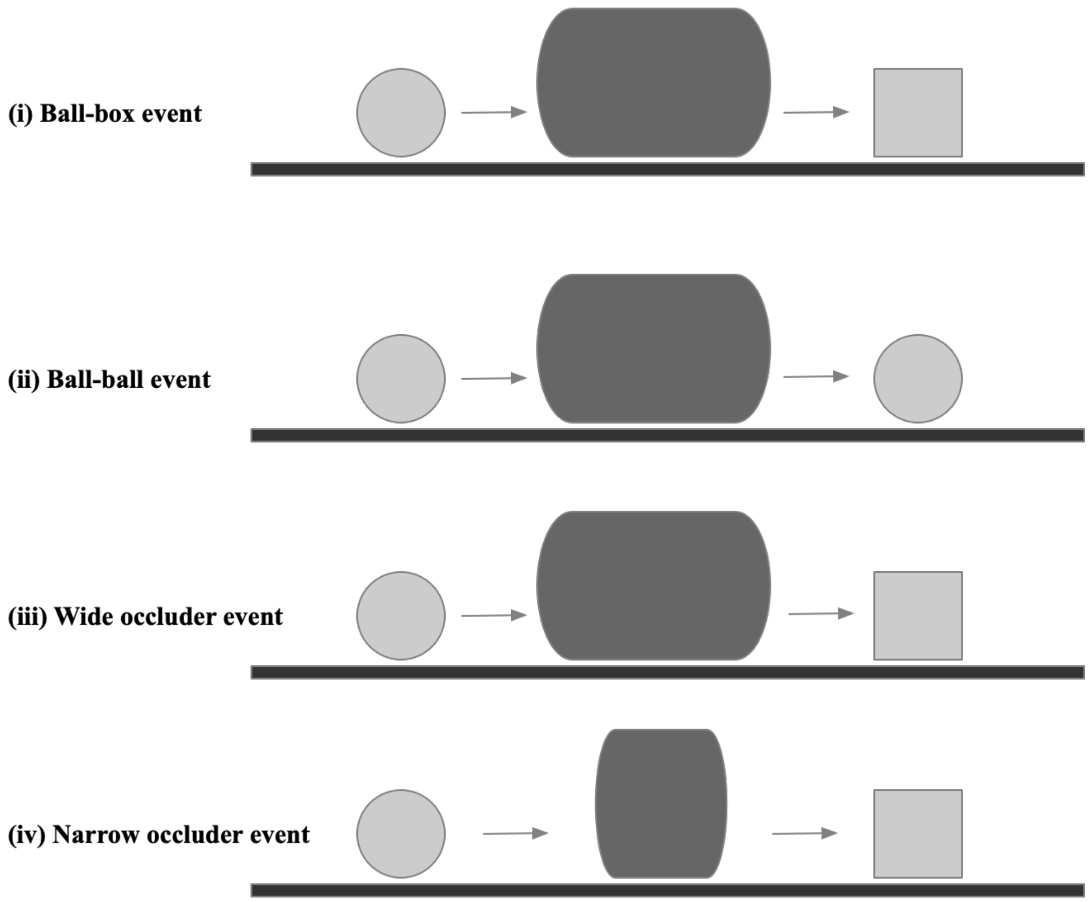

Figure 14 (i) ball-box event: a ball rolled from the left to the occluder and a box emerged from the right; (ii) ball-ball event: a ball rolled from the left to the occlude and a box emerged from the right; (iii) wide occlude event, similar as (i) with a wide occluder; (iv) narrow occlude event, similar as (i) with a narrow occluder.

T. Wilcox and Baillargeon (1998a) has tested 9.5 and 11.5 months old infants for what properties infants used for comparison. Two set of conditions in habituation created (Figure 14): i) a ball-box event with a ball rolling from left towards the occluder and emerged from the right as a box; ii) a ball-ball event with a ball rolling from left towards the occluder and emerged from the right as a ball. In test phase of both events, the screen was removed to reveal that a ball was behind it. Results suggested that 9.5 months old infants did not have a significant preferential looking when the ball-box event was finally revealed with a ball only behind the occluder. Contrarily, 11.5 months old infants looked significantly longer at the ball-box event.

Another similar experiment replicating the ball-box event with 9.5 and 7.5 months old infants intends to test the effect of the width of the occluder. Four tests were conducted with a wide screen and a narrow screen against the size of the ball and box: i) wide screen with large ball and box; ii) narrow screen with large ball and box; iii) wide screen with very small ball and box; iv) narrow screen with very small and box. Screens used in i) and ii) was just wide enough to hide a ball or a box while the screens in iii) and iv) could cover both ball and box perfectly. Results of test i) to iv) suggested that both age groups looked longer at the narrow screen with large ball and box.

T. Wilcox and Baillargeon (1998b) further tested the experiment with a group of 4.5-month-old infants with similar setting but the screen was decorated with stars. The experiment of ball-ball and ball-box events were replicated from T. Wilcox and Baillargeon (1998a) without removing the screen. 4.5-month-olds were found looking longer at the event with narrow-screen in a ball-box condition, like 7.5-month-olds. Thus, the research concludes that infants as young as 4.5 months old could compare properties in motion.

Teresa Wilcox (1999) repeated the procedures to examine four factors, shape, size, pattern, or colour, amongst 4.5 – 11.5 months of age and found that 4.5-month-olds could respond to the change by shape and size differences of the screens; 7.5-month-olds onwards could response to the difference of pattern; they respond to the colour difference only after 11.5 months of age.

It has been further examined that infants would prefer pattern than colour. Ng, Baillargeon, and Wilcox (2007) have also examined the colour effect amongst 8.5-month-old infants; infants were habituated to a setting with a screen and a cylinder which was moved by a hand to hide behind the screen. In the pattern event, the cylinder is in green with yellow dots initially. Then the hand would bring out a similar green cylinder but with yellow stripes. In the colour event, the hand would bring out another similar green cylinder but with red dots. The test subsequently proceeded with the hand bringing out the original yellow-dots cylinder. Finally, the screen lowered revealing a second transparent screen with no cylinder hidden behind. Results suggest that 8.5-month-olds did not respond differently to the colour event but only to the pattern difference. This result agrees with Kellman and Spelke (1983) that colour comparison developed in very late stage.

Discussion

Age of Infants (months old)	Responsive to difference of
4.5	Shape and size
7.5	Pattern
8.5	Pattern
11.5	Colour

Table 2 Summary of Change Responses

This suggested that when the trajectory path is not being tracked with wide occluder (Experiment 3), the reappearance of the object was not being compared with the object before it is occluded. With a wide occluder, the reappeared object is being tracked as another object. Only when the occluder is narrow enough, the object would be represented to track its motion with the occluder and therefore caused a surprise if the visible perception received is different from the representation when the object reappeared. As summarise in table 2, the surprise is caused by size and shape of the object initially but not pattern and colour.

Size and shape differences can be interpreted as a disappearance of certain properties at a computed position supported by coordinate system and the representation of the object. A re-encoding of the object can be updated if size and shape changed to produce proper coordination prediction in the case of a narrow occluder and therefore took more times to process. In a wide occluder condition, a new representation is created from perceivable signals for representation. But is the representation updated to the same object? Or it is a creation of a new object?

Experiment 8 – Two fully hidden objects

(a) Two screen test **(b) One screen test**

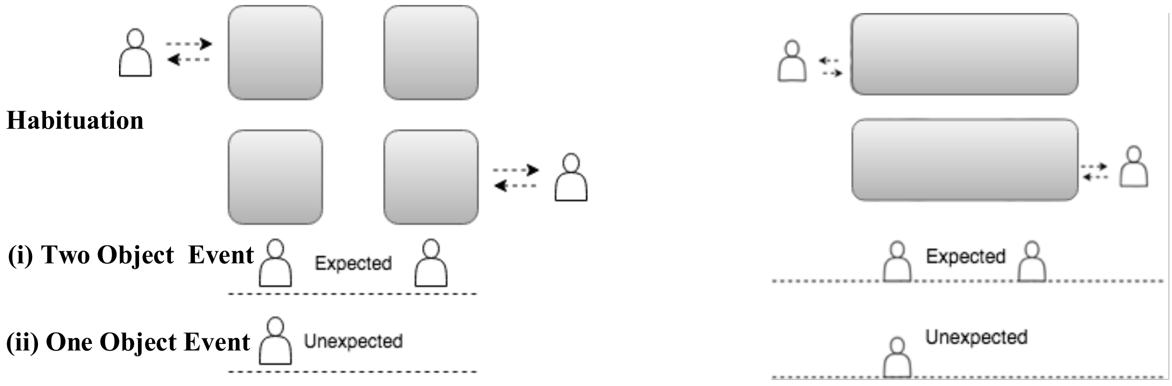

Figure 15 Objects were brought out from two screen respectively. In the expected event, the screens removed revealed two objects; in the unexpected one, only one object was seen.

Xu and Carey (1996) established a test with an occlusion event with two screens and two toys against 10-month-olds (Figure 15, (a)). During habituation, one toy emerged from left and hide behind the left occluder. Another toy emerged from right and hide behind the right occluder. Then in testing phase, both occluders were removed and revealed with either (i) two objects or (ii) one object (a). If the object was moving continuously and was visible between the screens in habituation, 10-month-olds looked longer at (i) two object event and expect it as one object. If the object was moving discontinuously in habituation (i.e. invisible between the screens), infants looked surprised to (ii) one object event and perceived it as two object.

The experiment was replicated replacing two occluders with one occluder (b) using a pair of different toys, a duck and a ball with same age group. In habituation, the toys appeared from either left or right from the occluder similar to the first experiment (a). Infants were found looking longer when two objects were revealed the screen.

One more test conducted with the same procedures but before habituation, both objects were brought out from the occluder to each side and stayed for a few seconds for infants to perceive. Under such condition, infants looked indifferently towards one or two objects. Subsequently, a similar experiment conducted for 12-month-olds, they looked longer at one objects after the removal of the occluder. Only until 12-month-olds, a different object appeared caused an expectation of two objects.

Discussion

New representation does not necessary induce a new object creation. Under a wide occluder, the representation of a different object would be re-created from the perceived toy reappeared. The representation is obviously updated back to refer to the same object. Therefore, it is a surprise if two objects were revealed. Only if the object was being tracked and represented when it is hidden, then the change of the properties of an object would create a new object with the new representation.

Experiment 9 – Gestalt relationship

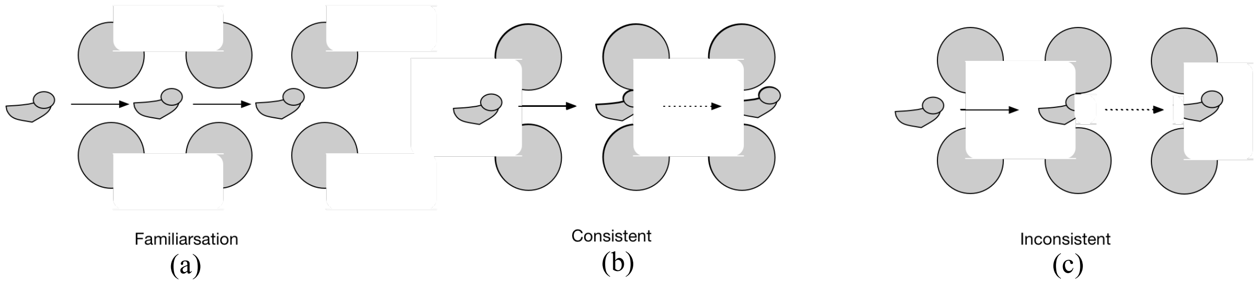

Figure 16 Illusionary occlusion event by Kanizsa

Csibra (2001) investigated 5- and 8-month-old infants to perceive continuity in illusionary occlusion event as in Figure 16. Infants were habituated to 6 circles with a quarter of it removed to form a pacman figure and a horizontally moving duck swimming between top 3 and bottom 3 of the pacman figures (Figure 16, a). Then they received consistent test (b) with openings of four of the pacman figures facing each other to form an illusory square; the duck swam behind the illusionary square invisibly and only visible outside the square. The inconsistent test (c) was also conducted with that the duck was seen swimming across the illusory square but invisible outside the square. Both age groups would either test with consistent test (b) first then inconsistent test (c), or inconsistent test (c) first then received consistent test (b). Results show that 8-month-old infants look longer at the inconsistent test (c) disregard of which test was conducted first while 5-month-olds looked longer with the first test disregard of whether it is consistent (b) or inconsistent (c).

Discussion

The illusionary square could not be interpreted as an occluder by 5 months old infants. 5-month-olds were surprised by the square itself provided the visibility of the swimming path of the duck was the same in consistent (b) and inconsistent test (c). Once they are familiarised with such an illusionary square then they are no longer surprised with inconsistent test (c). Contrarily, 8-month-old is able to represent an illusionary square as an occluder. Therefore, when the duck appeared in front of it, the inputs of the object should still be supplying from vision but it disappeared. This caused the surprise disregard of the test sequence of consistent (b) or inconsistent test (c).

This further implies that by 8-month-old, depth can already be used to support what is seen and what is not seen. This must be explained through a comparison of the depth of the object and the depth of the illusionary square. But how could the measurement of depth be learned with a meaning? Some more studies will be provided in next Chapter.

Experiment 10 – Containment with a lid

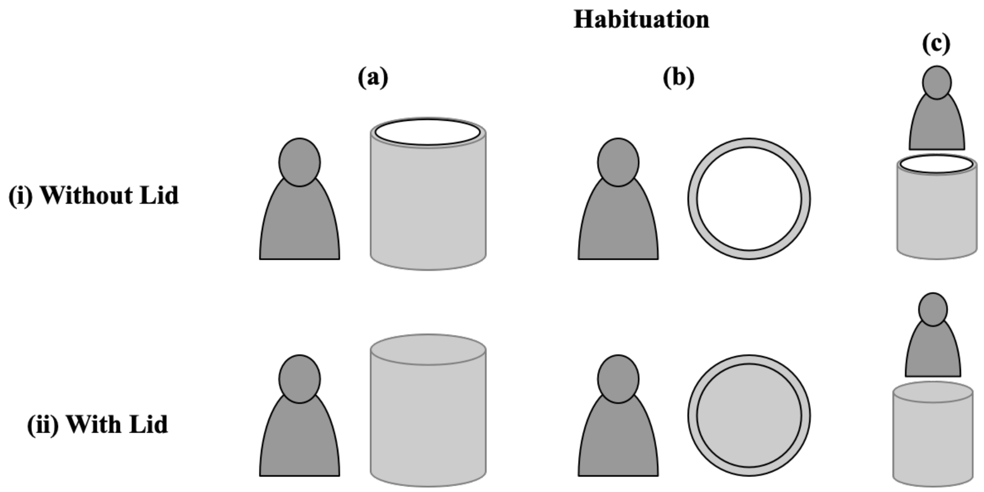

Figure 17 Figures adopted from S. J. Hespos and Baillargeon (2001b) and modified.
Habituation: (a) - object and the container line up side by side;
(b) – the top of the container was shown to infants;
(c) – the object was hang above the container but not lowering inside.

S. J. Hespos and Baillargeon (2001b) suggests that infants as young as 2-month-old could perceive objects in containment events. Infants was firstly habituated to the containers with and without lid as shown in Figure 17 without lowering the object inside the container.

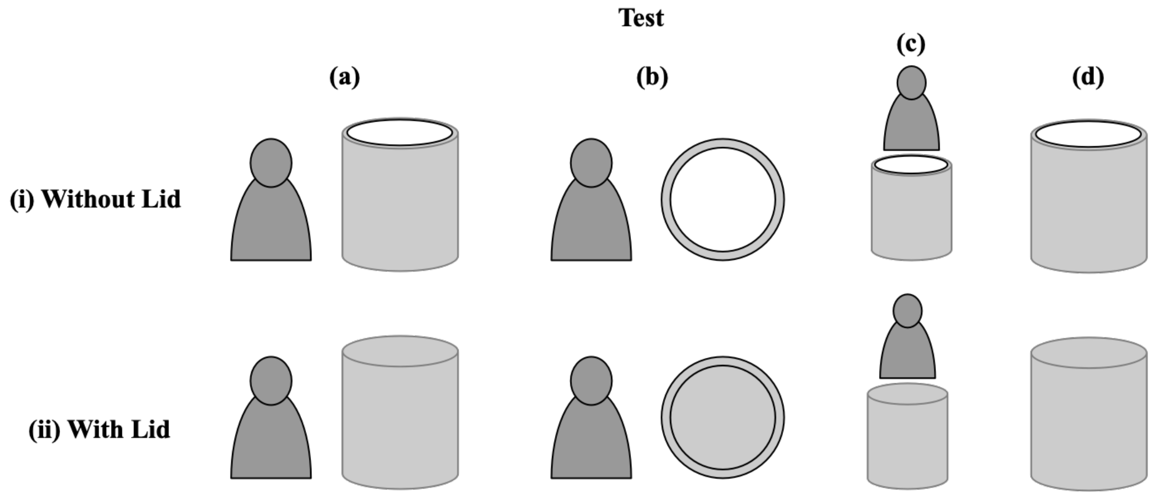

Figure 18 [S. J. Hespos and Baillargeon (2001b)]
The habituation step was repeated in the test but the object lowered into the container from (a) to (c).
then the object lowered into the container and hidden fully in both (i) and (ii)

They conducted two experiments that first one is about lowering an object into a container (i) without a lid and (ii) with a lid as in Figure 18. The object was lowered inside the container (d) and became full invisible. Infants were surprised when the object lowered into a container with a lid and disappeared.

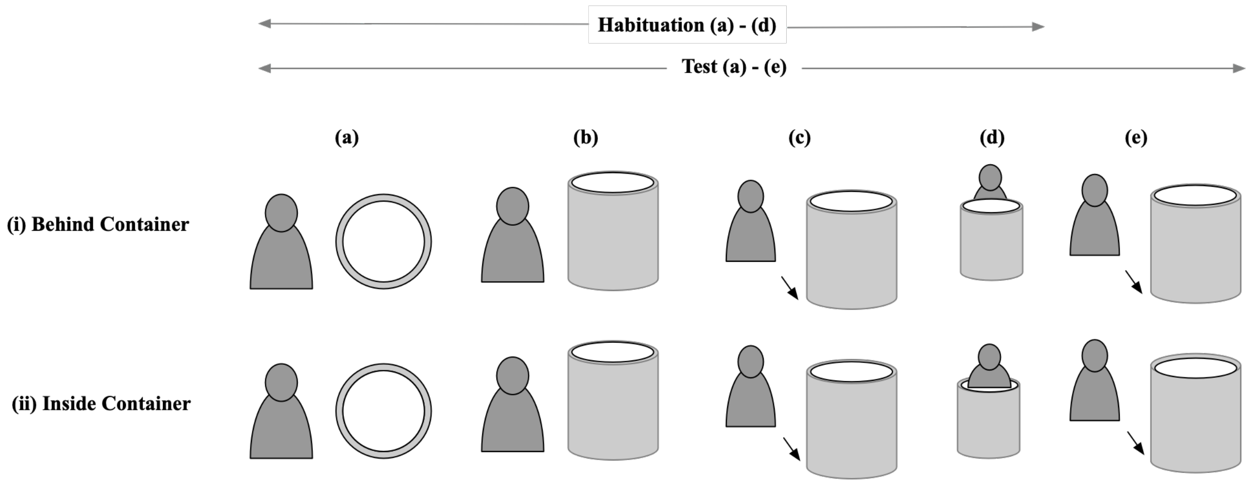

Figure 19 figure adopted from Hespos and Baillargeon (2001b).
Habituation steps from (a) show the top of the container to infants;
(b) line up the object and container side by side;
(c) move forward the container;
(d) lower the object behind (i) and inside (ii) the container;
Test steps repeated (a) to (d) and then (e) the object the container moved forward revealing an object behind;

Subsequently, a second experiment conducted by S. J. Hespos and Baillargeon (2001b) with 3-months-old that infants were habituated to lowering an object inside or behind a container without revealing where the object is Figure 19. Then infants were tested against both (i) and (ii) by removing the container away; the object was exposed to infants without moving together with container for both (i) and (ii). Infants looked longer at the left-behind object which was placed inside to the container.

Discussion

2-month-olds perceive a contact of surfaces is an end point for a motion which would cause the coordinate of two surfaces to be overlapping each other, just like impenetrable. A surprise of the object being hidden inside a container with a lid is actually an invalidated condition for the change of motion. The object must remain its visibility status after the contact.

When the object is lowered into the container, a contact of surface implies that the tracking point of the surface was now "inside" the container. The tracking point of the object is now represented with the container surface. This cannot be a comparison of the shape of the container with the hidden object as 3-month-old cannot compare representation with a visible surface (Experiment 5).

An object is being contained in a container means that the position for the object inside a container is represented by one coordination set only, i.e. same as the container. The coordination system maintains and applies changes that are applicable to both representation of the object and the container. Therefore, a positional change of the container will induce an expectation of coordination change of the object.

Experiment 11 – Containment versus occlusion

Some experiments seem to suggest that infants could respond to width difference between objects. S. H. Wang, Baillargeon, and Brueckner (2004) have examined 4-month-old infants against placing an object in or behind a container. Infants were found looking longer when a wider object is inside or behind a narrower container. This experiment did not habituate participating infants and they could still demonstrate a different response. Without habituation, infants showed that they could compare the width of the object and container and respond to the disappearance of a wider object. If the disappearance of a wider object is a more significant change to them while the narrower one is not, this means infants can compare a group of properties against the other group of properties one-dimensionally by 4-month-old.

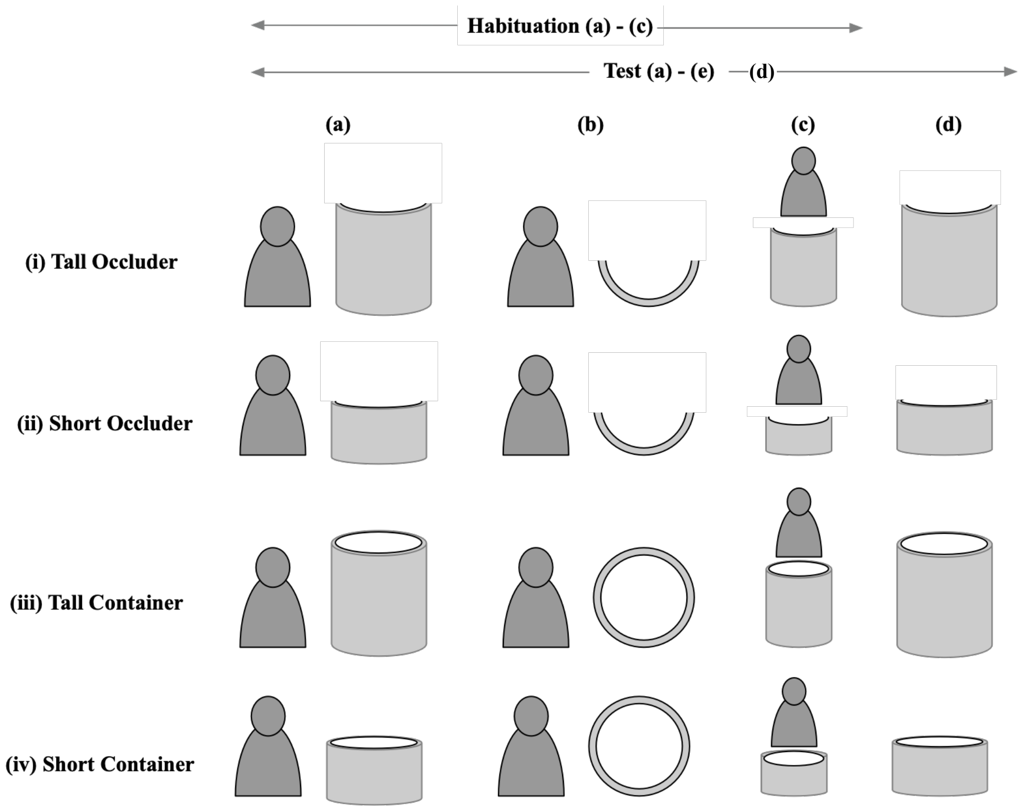

Figure 20 Figure adopted from Hespos and Baillargeon (2001a) and modified.
Habituation (a) – (c): (a) an object lined up with either (i) a tall occluder, (ii) a short occluder, (iii) a tall container and (iv) a short container; (b) the top view of the occluder or the container was shown to infants; (c) the object was hanged above the occluder or container without lowering down.
Test (a) – (d): Steps (a) to (c) were repeated in test phase and (d) the object was then lowered down and became fully invisible in all (i) to (iv) events.

Another experiment suggested that 4.5-month-old infants attend to the changes of height difference. S. J. Hespos and Baillargeon (2001a) habituated infants by showing them (b) the top view (b) of the container for container tests (Figure 20, (iii) and (iv)) and (b) the cross section of an arc of the occluder for occluder tests (Figure 20, (i) and (ii)).

Subsequently, the object was being lifted and (c) hanged above the container or the occluder for 1 second and then return to the left position without lowering them inside the container or behind the occluder. Then in container tests, (d) a tall object was lowered into a tall container and a short container. Similarly in occluder test, (d) the tall object was lowered behind the occluder. After the object was lowered inside the containers or behind the occluder, it became fully hidden in both tall and short container or occluder. Results show that 4.5-month-old looked longer only to the short occluder test but indifferently to both container tests.

The same experiment conducted with modification that the container was moved forward towards the infants before lifting the object in habituation and before lowering the object inside the container. Since the object was not moved forward, it was behind the container when it was lifted and lowered similar to Figure 19 (i) (d). After the object was fully hidden, the container was moved backward to the original position. Results showed that infants looked longer when the object was fully hidden behind the short container.

The first experiment of lowering the object into the container (not behind the container) was repeated for 5.5-, 6.5-, and 7.5-month-old infants. It was found that only 7.5 months old infants looked longer at the short container event with the tall object fully hidden.

Discussion

4-month-old responses to width difference can be due to the disappeared visible perceptual signals for longer object when it is lowered into a narrower container or occluder. The wider object must remain on top due to a contact of the wall of the container and the object, i.e. the end point of the motion. But the coordinate continued to change and caused the surprise.

For the height experiment, the container can be viewed with an inside. However, height comparison is indeed a comparison of difference of maximum and minimum of Y coordinates from a surface with X coordinate endpoints. The container was however in a cylindrical shape and could not be extracted with a surface two-dimensionally for comparison. The cylindrical shape must be learnt with computation of a volume three-dimensionally. Therefore, before such a learning happened, the comparison of the container and the object cannot be performed. Explanation of containment has to firstly make sense out of an "inside". "Inside" is indeed not a two-dimensionally measurement that could be related by eyes (only responding to X and Y dimensional changes). It involves an understanding of the meaning of Z dimension (i.e. depth) contributing to a volume. It was until rough 7.5 to 8 months old, infants understand depth. If infants do not understand the meaning of depth, how could infants understand occlusion?

Occlusion also makes use of depth, however, it only involves comparison of surfaces between the occluder and the object. Given the understanding of an end point for a motion at a contact event, a continuation of motion implies a "behind". But this "behind" does not involve a construction of a volume for comparison. So, occlusion can only be explained as "behind" or not. That is the perceptual signals received are different from the represented object linked and computed by the coordinate system. It is just a seen inputs (occluder) and unseen representation (hidden object).

General Discussion

Table 3 Summary of Interpretation of Empirical Findings

Age	**Interpretation**
A few hours	• Attend to highly contrasted properties from part of an object
2-month-old	• track properties moved and changed together with observable discrepancy • can represent hidden motion path as predictive computation from recent positional changes • tend not to track the motion path when the moving object is invisible with wide occluder • can reason a contact between two visible surfaces as an end point of motion
3.5-month-old	• can represent and compare a partially hidden object with a visible occluder
4-month-old	• tend to connect properties that moved in the same positional changes • can validate the dimension of a hidden object against another visible object • can compare visibility of objects using surfaces of the objects as seen and unseen
8-month-old	• can understand representation of illusionary surfaces for comparison

	• can use depth to validate representations
12-month-old	• more than one representation can be created to represent various hidden objects differentiated by shape, size, pattern and colours even with the same positional changes

Summarising empirical findings, i) motion is essential to group properties together for tracking of positional changes; ii) tracking facilitates a formation of object in terms of property groups; iii) tracking of motion requires a coordinate systems minimally; iv) occlusion event in motion must engage representation system to track a hidden object.

Perceptual learning was initially led by motion and tracking of positional changes with a coordinate system. There are two factors to consider: i) change of motion; ii) change of visibility. A contact between two surfaces implies an end point for a motion. If the motion continues to execute, the representation of the object would be used to track the motion without interference by the signals of stationary visible barrier, i.e. the occluder. Under this circumstance, the constraint of being able to represent the hidden object from last snapshot is very high such that the representation can only be exploited for a narrow occluder in early stage. If the occluder becomes wider, early infancy tends to jump to the exit of the occluder to continue track the object instead of representing the object and track it as a continuous motion path. Therefore, the width of occluder determines the availability of the representation of the object for comparison. With the representation and a tendency to connect properties moved together, older infants are able to raise question to validate the "expected" connection. The question could be answered with later updates from a snapshot even with observable discrepancy in time.

Seen and unseen of an object refer to different inputs. Stationary occluder is supplying a direct perceptual inputs all the times while an object in motion must have to be represented. Representation created for motion tracking enables comparison between the representation and the visible inputs. This supports the behaviour of occlusion events such that the representation of an object used in a motion can be compared against X and Y dimensions with other inputs. Comparison involving surfaces is available. Up to this stage, even if Z dimension is perceived, the subject does not seem to understand the meaning for comparison. The coordinate system only assigns the coordinates to the representation and eye scanning only gives a meaning to two-dimensional comparison. This could be supported with the cylindrical container that a surface cannot be formed to compare two-dimensional coordinates between the representational object that is placed inside the container to the visible object such as height. It can be compared if it is behind an arc of surface.

The key turning point happens at around 7.5 to 8 months old that infants seem to be able to understand depth. Before this age, following a motion is only a two-dimensional coordinate changes that could be understood with a meaning of how much the eye shifts. However, depth perceived cannot be converted by units of eye shifts (e.g. how much force is exerted to keep the coordinates of stimuli in the center position).

The advantage of the motion tracking process with coordinate and representation system is that it does not stop infants from processing the object with missing inputs. They can make use of the internal representation as a reference for further processing and chain up reactions. Another advantage is to make use of previous experience in form of memory and invoke previous learnings for faster reaction. This could constitute a basic unit for a mind process that goes beyond direct perception for symbolic reasoning, i.e. allowing rich internal processing with a continuous supply of inputs without reliance on direct signals.

Given the experiment in Kellman and Spelke (1983), it is really hard to tell whether infants understand occlusion in habituation phase without a full sight of the occluded rod. Only when infants see a complete rod in test phase, they could associate the similarity of the complete rod with the (represented) occluded rod in motion and apply the same tracking responses, then that could contribute to a retrospective update of the invisible part of the occluded rod and relate the two display with a common (represented) rod and a description of the positional change and other episodic information. This could lead to a learning of occlusion with respect to the experimental scenario. It could also be possible that infants have already seen some similar rod before and can compare the experimental scenario with their own previous experience. However, unseen is unseen. It should still form a question to be answered as a form of expectation and causes more response to find out the answer. In such case, it might be the most primitive drive of exploration in a form of cognitive question.

5 How can a basic representation system develop further?

If one processed physical objects perception with the work of static objects only, the work cannot adequately explain the behaviour of why infants respond to changing objects. For example, psychologists suggested that infants do not just respond to static object but also objects with changing features such as self-propelled toys (Luo, Kaufman, & Baillargeon, Cognitive psychology/2009). If infants viewed objects as a list of properties, how could infants relate that list of properties that vary? Animals and human beings are featured with many changing properties. Can infants recognise them? How could infants relate them as the same object?

New-borns were found with a strong attachment to their parents, especially their mothers (Bowlby, 1958, 1969). Similarly, other species are also found to have active interactions and strong bonding in early stage with their caretakers (Hayashi & Matsuzawa, 2017; Pissonnier, Thiery, Fabre-Nys, Poindron, & Keverne, 1985; Peter Szenczi, Banszegi, Urrutia, Hudson, & Farago, 2016). What could infants know about their caretakers? Through an examination of what could be recognised from their mothers, this section explores insights to represent animated objects.

5.1 What Do Infants Recognise from Mothers?

Psychologists suggest that new-borns could recognise their mothers. Bushneil, Sai, and Mullin (1989) found that 2-day-olds recognised their mother. Their experiment sampled female adults with similar hair colour and facial complexion but different hair-style or facial shape to make the female adults "broadly comparable"; these female adults were sprayed with a strong air-freshener to avoid olfactory confinement and were asked to sit behind a screen exposing their faces only. Results suggested that visual perception of the face alone could make a difference of a female stranger from a "mother". The question is what does it matter for recognition?

Mother-infant bonding is said to be strongly formed in a few days after birth in terms of recognition. Bushnell (2001) observed the relation of time spent with mothers and preference for their mothers amongst new-borns who were 2 to 7 hours old in hospital; the research observed that infants gradually got used to the facial features of their mothers in first few days. Bushnell (2001) further examined 2 to 4 days old infants in hospital; the experiment investigated the relation of mother recognition with time. Result suggested that a longer duration of more than 15 minutes is required to remove the memory of mother and the memory is already very stable in a few days.

The above empirical finding proposed that even very young infants could respond to their mothers. However, their mothers could not be invariantly perceived and might not be moving in conformance with physical rules. This recognition ability seems to fail the computational models that predicted a motion path of static objects by training, as well as representing an object with list of exact features such as sizes, shapes and colour. In just a few days' or even in a few hours' exposure, the bonding between infant and mother is strongly formed in very early infancy. The strong bond formed cannot be explained solely as a consequence of intensive exposures. Contrarily, awakened time for new-borns is very limited. A research has reported that only 16.2% of time of new-borns is awakened in first two days (Freudigman & Thoman, 1993), and 15% of time in 2- to 3-day-olds (Sadeh, Acebo, Seifer, Aytur, & Carskadon, 1995). With limited exposure time in first few days of a new life, what could infants exploit to recognize their mothers?

5.1.1 Partial Feature Recognition

Experiment 14 – Face Recognition

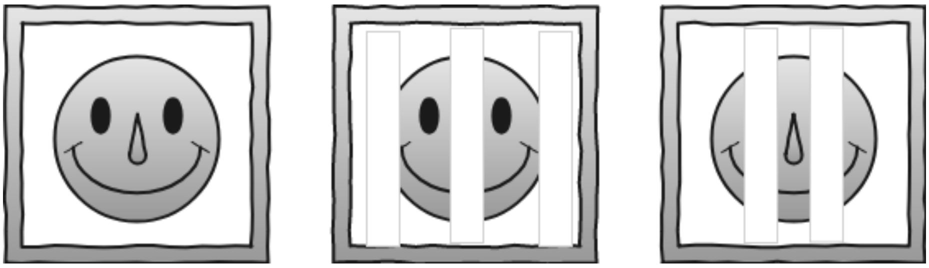

Figure 21 Photographs of Caucasian females with vertical white strips

Infants could not distinguish all the detail features of their mothers with stranger females. Gava, Valenza, Turati, and Schonen (2008) found that 1 to 2 days old infants do not read the whole face for recognition. The first experiment presented infants with upright or upside-down grayscale photographs of Caucasian females. The photographs consisted of a pair of illustrations, one of them showed eyes and the other showed mouth and nose. They found that infants only show preference towards upright face if eyes were visible but look indifferently to upright or upside-down photos if eyes were occluded. This experiment proved that infants respond to certain features only (referred by the research as "highly salient features" such as eyes). The second experiment extended this finding by habituating infants with a non-occluded grayscale photograph of a female and testing them with occluded photos of a familiar face from habituation and a novel face that: i) showed nose and mouth only ("low salient features"), and ii) showed hair and eyes only ("high salient features"). The other untested features were occluded by vertical bar as in Figure 21. This experiment intended to test if infants were able to recognise the same person when salient features were being occluded with vertical strips. This study further proved that infants were able to recognise the same photograph if eyes were not blocked but not nose and mouth.

Even if a partial view of the face in a different angle was sighted, infants could still respond to familiar faces. Turati, Bulf, and Simion (2008) showed that infants are familiar with both frontal view and partial view of mother's face. They habituated 1- to 3-day-old infants with: i) ½ face picture and tested with full-face poses; ii) full-face picture and tested with ¾ poses, iii) ½ face picture and tested with ¾ poses, iv) full-face picture and ¾ poses and tested against each other. Their results stated that infants are able to respond to faces shown in full and ¾ poses. They concluded that turning of a face up to ¾ faces being visible could still be recognised by infants.

The above finding is however performed under static image switching between snapshots in habituation and test phase but did not engage any motion; further investigation about turning a face was being explored with 1 to 3 days old infants in Bulf and Turati (2010). They ran experiments with static view and smoothly rotating female faces in both orderly manner and random orders; infants were habituated with full face of 30º, 60º, 90º profile poses without hairline in three different manners: i) smoothly from left face to right face, ii) static snapshots in order of i), iii) random sequence of -60º, -30º, full face, +30º, +90º. Subsequently, infants would be tested with familiar and novel face in profile pose. In i), infants were able to recognized familiar face much better than ii) and iii) the worst. This experiment has demonstrated that, similar to static objects, smooth motion facilitates recognition of an object without a full view of the object.

Discussion

These experiments have shown importance of distinguishable features in recognition that are more crucial than comparison of a list of perceptual facts. Echoed Experiment 3, infants can follow the same object with observable discrepancy. For animated properties such as eyes, it could be just a recognition of a particular snapshot of an eye. Mouth and eyes are considered to be more contrasting than nose with motion that might catch the attention to follow and encode them to produce rich representations. They could be used as the key properties for tracking or other response. Less salient properties might be either learned as a weak representation or ignored that could not be used as a key to distinguish different faces. When the snapshot of a face changed, the recent representation could be updated due to a change at that moment and therefore can still be followed as the same object.

Experiment 15 – Spatial Relations

The above findings have not fully isolated the eyes as the only stimuli but also tested with the outer hairline being shown partially; possible confounding was possible to interfere with the conclusion. Seeking further evidence, another experiment conducted by Turati, Macchi Cassia, Simion, and Leo (2006) investigated if inner or outer facial features interfered with infants' recognition. 1 to 3 days old infants were habituated with grayscale photographs of females and were presented with: i) full face, ii) inner facial features only and iii) outer facial with blank face photos during test phase. Results showed that infants could distinguish the face with inner or outer facial features only. This result does not just agree with Gava et al. (2008) about inner saliency as a major crucial factor for recognition but extended further that outer features such as hairline and hair style could also be the one.

Turati, Di Giorgio, Bardi, and Simion (2010) has further examined the recognition of first halves of inner facial features with respect to second halves alignment. 3-month-olds were habituated to first halved face without outer features and then were tested with an aligned or misaligned second halved. Results suggested that infants, same as adults, focused on eye area of the face even with aligned second halved. These experiments have indicated that position alignment of facial features is one of the attributes to recognize an object.

The outer features of both hairline and hair styles have been found crucial in recognition with two separate experiments. Pascalis, de Schonen, Morton, Deruelle, and Fabre-Grenet (1995) have raised evidences to support that infants could recognise the difference of hairline. 4 days old infants were presented with mother and stranger female wearing scarf under the chin; the contrast of the scarf and face were set to be minimized to simulate a change of outer contour. Results showed that infants did not recognize their mothers if hairline contour changed. Bushnell (2003) examined 2 days old infants for hair style factor and suggested that it is another crucial factor. They exploited the same mother-stranger discrimination task with grayscale photos but changing the hair style with a wig. The wig was placed on the photo either covering the original hair style or inverted below the chin exposing the original hair style. This experiment figured out that infants significantly looked longer at the picture of mother with inverted wig condition but not normal wig one. These two findings added outer features into the crucial mother recognition factors.

Both Turati et al. (2006) and Gava et al. (2008)'s experiment actually involved more than featural information but also a recognition of upright position of a face that could have been contributed by shapes, feature distance and positions. Infants seems to know the position of facial features. Leo and Simion (2009a) has tested the overall face position and also particular feature's position with 1-to-3 days old infants and found that they could recognize familiar face only if the position was normal. First experiment has habituated infants with either upright or inverted eyes of a full face in grayscale photograph and then tested with novelty and familiarity; the novelty scores stated that infants are able to familiarise with the face if they were habituated to a "normal" face instead of a strangely arranged face. This could support that new-borns recognise a face by the spatial relation of facial features with a human face. Furthermore, they also tested 1-to-2 days old infants against the spatial arrangement of overall face. Similar procedures were applied and infants were habituated to either upright face or inverted face with full outer and inner features in grayscale photographs; both orientations has either been presented with unaltered or thatcherized eyes. Results showed that only if the face was presented in upright position in habituation, infants would be able to recognise if the face is altered. The overall spatial configuration of the face being learnt has a determining influence for recognition.

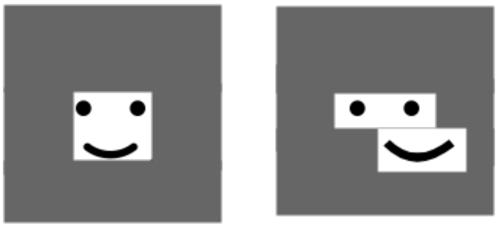

Figure 22 Experiment tested with unaligned first and second half of a face (Turati et al., 2010)

Learning of the features of a face could be strengthened overtime even if it was disoriented. Top and bottom face alignment experiment in Turati et al. (2010) has also found a different response from 2-day-olds and 3-month-olds; as in Figure 22, 3-month-olds looked longer at familiar face when it was not aligned with second halved while 2-day-olds didn't.

Figure 23 Two-toned mooney face in Leo and Simion (2009b)

Compared with another experiment by Leo and Simion (2009b) that tested 1-to-3 days old infants' novelty response against two-toned mooney photographs. The mooney face only drew with some black-and-white shapes indicating proper location of some facial features without details as shown in Figure 23. Infants prefer upright mooney face photographs against inverted one and a non-face object. Latter work showed that infants are able to detect the overall spatial arrangement of a face. Instead of suggesting that infants only know a face's spatial configuration in later stage, it might be worthwhile to consider that 2-days old could respond to spatial relations of facial features with minimal details of properties at first and developed into detail recognition of the whole face in later stage.

Discussion

From these evidences, they suggested that infants do not exploit all the featural details of a face to refer to the same object. The validation of a matching for recognition only happens around a couple of properties only. These properties seem could be identified by itself alone and also part of another object (e.g. the face) that "contains" the properties. This relationship can be described by a spatial relation supported by the coordination system.

Properties can all be represented by the relative coordinates describing the spatial relation of the properties with other properties that occupy a specific amount of the space and describe as a form of relative distance to other facial properties. This does not just imply that the coordinate system works in a relative magnitude but also relative to other objects at properties level, in a loose manner, in order to support different degree of poses and dynamic changes of an object.

The level of detail of properties could be the granularity of distinguishability. First determining factor is the boundary of a shape that is supported by discontinuity of a shade (or brightness). For example, the outer boundary of a face is a connection point of a face and hairline. A face contains eyes, nose and mouth. The other factor could be motion. The common characteristic of eyes and mouth is the expressive mobility on a face that allows them to be tracked and followed by infants. A meaningful grouping of properties could be by connectedness of shades first (contrast of difference between static properties) and then further grouped by motion like static objects reviewed in chapter 3. Mouth and eyes can move on their own but at the same time they cannot be left outside a face if a head was turning. A description of the spatial configuration (i.e. relative coordinates of a properties group connecting other properties) should be able to be represented.

5.1.2 Experiential Learning

Experiment 16 – Construction of More Knowledge

Learning capability seems to be greatly influenced by similarity of the examples in daily exposure. Kelly et al. (2009) showed that race difference in human could influence the familiarisation of new faces for the same race. 3, 6 and 9 months old of Han Chinese infants were habituated to a combination of photographs of African, Chinese and Caucasian adults before a test with 1 familiar and 1 novel faces. Habituation figures showed that 9-month-olds were fastest to be habituated and 3-month-olds were slowest. Results showed that 6 and 9 months old Chinese infants have significant novelty preference to Chinese stranger face but not the other two races, but 3-month-olds demonstrate novelty preference to all races. These suggest that the more repetitive exposure to similar features, the more biased the recognition capability could be derived towards objects with similar features.

More evidences could be found from holding-side influences of new-borns by their mother that give rise to a perceptual bias for face recognition. Vervloed, Hendriks, and van den Eijnde (2011) investigated universities students for their biasness in relation to the handedness of their mothers who bottle feed their babies. Students were presented to greyscale photographs with a facial expression formed by left halved and right halved of different faces; one set of photographs held two different emotions with left or right halved of a face while the other set held different genders of similar faces. Then students would describe the face with a closest emotion or determine the gender. Results showed that all students with a left-holding mother show a left-bias towards both emotion and gender facial stimulus. Even those with a right-holding mother shown with left-bias tendency, it is significantly reduced.

Their hypothesis was further verified in the work of Hendriks, van Rijswijk, and Omtzigt (2011); they recruited mothers to simulate bottle-feeding a doll with inbuilt camera. Results showed that if a doll was being held on right arm, mother's face was less detectable from the doll compared with the other side. From the recorded images of holding the doll in right arm, it is found that left half of the face was less visible by lookup action of the mother while right half of the face was less visible when they are looking at the doll. They suggested that left halved of the face considered to be more expressive, holding a baby in right arm could leads to less exposure to the overall facial expression. These findings explained the theory that left-bias tendency reduction in the students is due to suboptimal facial expression exposure if they were being right-held by their mother.

Discussion

These works have illustrated exposure of previous experience has determining distraction towards future representation construction in a way that new knowledge tends to be explored and built on previously familiarised properties. The experiments have demonstrated a "biasness" of further construction of knowledge. This is different from favours. The biasness is influenced by previous perceptions of similar objects. The similarity could be referred as partial recognition. The partial recognition imposes a tendency to retrieve and search from previous knowledge to form a basic representation, then the representation could be used to build a new representation of an object with further updates. This could perform like a bias in terms of experiences.

5.1.3 Multi-modal Redundancy

Experiment 17 – Other sensory recognition

Infants were able to synthesize various sensations together that are collected from their mothers to form a representation. When visual features were salient enough, they would be able to just use visual inputs to identify their mother as in Bushneil et al. (1989). Is this alone adequate for the recognition?

With reference to the nature, it has been discovered that animals could also make use of multi-modal information to identify their mother such as smell and sounds. Péter Szenczi, Bánszegi, Urrutia, Faragó, and Hudson (2016) suggested that kitten is able to recognise its mother by the voice. Pissonnier et al. (1985) studied olfactory signals and found that they are significant to build the bonding with mother sheep. These might help to uniquely remember a mother in infancy.

Infants are able to recognise the voice of their mothers. Sai (2005) depicted an experiment with a few hours new-borns to determine if interaction and voice were the factors to recognize their mothers. First experiment with 2-to-4-hour-olds has shown infants images of their own mother and a similar female with their heads only but hiding the olfactory and voice features; result reconfirmed previous findings that infants are able to recognize their mothers even within a few hours after birth. Similar experiment conducted with 2-to-12-hour-olds but allowed interactions between mother and baby as well as implemented with vigorous control on mothers to avoid making any sounds to their babies after birth. Results showed that infants were not showing more fixation on their own mother. They also studied that voice has an effect of directing infants' attention by turning their heads often towards the stimuli. 6-hour-olds who were allowed to have voice experiences with mothers were compared with those who weren't; evidences suggested that those with voice experiences with mother turned their heads more often to mothers than to strangers. These results have illustrated that the influence of voice in new-borns. Infants with limited exposure to their mothers might not be able to recognise their mothers without voice.

Significant evidences have shown that infants who have been breast-feeding by mothers are able to build a preference towards the same odour. Macfarlane (1975) performed the test with new-borns in hospital that were breast fed by their mothers in 3 to 4 hours' time; the mothers were asked to use standard hospital breast pad. Then babies were tested against two pads hanging on top of them, one was from their own mother and another was a clean one. Results have illustrated 2-day-olds spent 51.8% of time turned to their own mother's breast pad; 6-day-olds spent 60.3% while 8-10 days spent 68.2%. The increasing trend has shown that new-borns are capable to distinguish their mother's olfactory cues.

A similar findings from Russell (1976) agreed with Macfarlane (1975); they recruited second day full-term breast-feeding babies from hospital wards and their mothers were asked to put on a breast pad for 3 hours before a test. In the test, infants were observed when presented with a clean moist pad, a stranger's pad and their own mother's pad. It is found that 2 days old infants did not respond to their own mothers' pad and the response increased a little bit with 2 weeks old and they would be able to recognise their own mother's odour only until 6 weeks old. These strongly supported that infants could recognise their own mother through olfactory cues. This is a learning resulted from intimate interaction with their own mothers as confirmed in the work of Cernoch and Porter (1985) that bottle feeding 2 weeks old infants do not exhibit the degree of responsiveness as breast feeding infants.

Some researchers suggested that olfactory cues cannot be exploited alone for mother recognition. Bushneil et al. (1989) has covered up the face of the mothers and strangers with a mask and tested if 54 hours old infants were able to show a preference to their mothers. Result demonstrated that they failed to recognise their mothers. While odour could have caught the attention of infants towards the stimuli, it is inadequate to distinguish mother without visually perceiving mother's face at least before 2 weeks old. This interpretation has not actually contradicted with Russell (1976) and Macfarlane (1975) as both illustrated 2-days-old has not been readily responding to their own mother's odour, not until 2-weeks-old. Therefore, olfactory cues are progressively learnt and strengthened through more intimate interaction between mothers and infants.

These visual, audio and olfactory works combined complex unimodal sensational experiences to claim a formation of multi-modal redundancy effect. Bahrick, Lickliter, and Castellanos (2013) tested 2-month-olds with synchronous audio-visual with a woman singing a song, asynchronous audio-visual with an unmatching mouth movement of the women, and unimodal visual speech without sound; infants were habituated to a woman first and then tested with a novel one. Results showed that infants could recover their visual quickly in synchronous audio-visual, slower in unimodal visual and slowest in asynchronous audio-visual situation.

Discussion

Different sensory inputs can aggregate together to recognise an object without reliance on a full visual representation. This implied that i) sensory inputs from an object could be related with different modal inputs; ii) when a sensory input is being received, the representation of the object could be associated with other sensory signals and raised a cognitive question for responses.

The relationship construction can be enabled by the synchronousness of the inputs. Synchronousness of sensory inputs could refer to the change of various sensory inputs versus the change of motion of an object visually. For example, to be able to relate visual input with smell, it could be the strength of smell against closeness of the object when the object is moving toward the subject. To relate sound with visual input, it could be the regularity of change of mouth against the change of sound frequency.

Once the association can be built, there is a "conversion" process to build a representation to resolve cognitive problems. When a sound of familiar object received, the visual signals being stored could be retrieved and raise a question to validate the visual representation. Subsequently, a response of turning the head towards the sound could be driven. This mechanism overcomes the limitation of each of the shortcoming of sensors such as a very short distance of vision in early stage and allows perceptual learning with different degree of maturity of different sensors.

5.1.4 General Discussion

Given the findings that infants have shown strong bonding with their caretakers especially their own mother (Ding, Xu, Wang, Li, & Wang, 2012), psychologists suggested that the bonding is a form of affective attachment (Bowlby, 1958, 1969). Disregard of whatever cause that draws infants to attach to their caretakers, to be able to identify a subject to attach is more than just being able to recognise a familiar face. It is a form of attachment that is highly adhesive to certain group of properties received from stimulation.

Behind this adhesiveness, empirical evidences implied a recognition of flexible groupings of properties from visual perception and their spatial relationships. A mobile properties group (eyes or mouth) could form a group of properties that can be representable in itself and facilitate recognition with spatial relationship with other group of properties. There is no guarantee of what could be seen in each snapshot and the flexibility creates dilemma for recognition. It is therefore an essential characteristic of caretaker that the representation must be created with richest properties base so that different combination of exposures could also lead to a recognizable object.

This adhesiveness can become stronger with multi-modal sensory inputs to overcome shortcoming of each sensors, respond to discover more properties and extend the coverage of recognition capability. The process of loosely coupling properties as an object is a general enablement (or process) across different sensation that does not depend on a particular sensor but supplementary to each other. They could recognize their mother's hairline, hair style, mouth and eyes visually; they could recognize the odour of their mothers with sufficient intimate interaction such as breast-feeding; they could make use of sounds to recognise familiar people. Results are not unanimous in all the infants of the experiments even if they are of the same age; this means that these sensational learning is not rigidly enforced as a mandatory comparison of a defined list of size, shapes, colour, location, etc. Instead, it is a dynamic grouping used in recognition that is capable to link multi-model sensory inputs together to represent the same object (Bahrick et al., 2013; Bushneil et al., 1989; Cernoch & Porter, 1985; Macfarlane, 1975; Russell, 1976; Sai, 2005).

Mobility of the subject affects visual learnings. To be able to form the attachment, infants could learn who are their caretakers by right orientation only initially. One of the essential findings about first 3 months is that infants started the learning of an object or a face with the spatial relation of the properties, i.e. the orientation. It echoes the finding that infants shifted from being sensitive to orientation change at 6-week-olds to angle differences at 14-week-olds (Leslie B. Cohen & Younger, 1984). It might support the negative result of novelty recognition in inverted faces that infants are not sensitive to novelty in disoriented photos in first few weeks. Only in a right orientation that they could respond to it. The possible explanation could be that they have not yet being too mobile that they are still learning to calibrate their own relative position in relation to the position and orientation of their immediate environment. This might cause a different orientation as a novel stimulation to them. The dynamic nature of learning the physical worlds as in infant learning is that both the agent and the objects are mobile. Infants themselves also move and change (growth) with more sensorimotor and exploratory power. The capability to adapt to these external or internal changes must consider the subject and are able to relate to the same object even the subject moved. Evidence has shown that infants might experience some confusion and result in recalibration in transition stage (Grzyb, Del Pobil, & Smith, 2012) when they acquired more capability such as walking (Karasik, Tamis-LeMonda, & Adolph, 2011). Consequently, the transitional stage of calibration of the distance between the subject and the immediate environment explores the meaning of relative coordination changes with the change of the subject.

Learning of novelty properties seem to be built from previously encountered properties. This characteristic determined certain behaviours they exhibited. For example, infants seem to be able to respond to commonality of the familiarized properties to learn novel features or objects by their difference. This tendency might bias what to be learnt and how quickly a novel object could be learnt. Evidences such as more effective recognition of novel face in own race (Kelly et al., 2009; Liu et al., 2015), or grown-ups biasness with respect to a repetitive patterned stimulation in infancy (Vervloed et al., 2011). It could have biased future responses derived out of their previous experiences (Hendriks et al., 2011).

5.2 What Does Depth Mean?

Computationally, infants' behaviours towards occlusion might not be able to represent depth fully. Representational processing of occlusion does not give rise to a full interpretation of Z dimension changes as occlusion experiments are just horizontal motion, i.e. changes of X dimension, and changes of visibility. Whether or not depth can be learnt with a meaning has to be examined with behaviours associated with depth changes. If a representation of a three-dimensional space cannot be properly formed, no representation could be properly formed for comparison with three-dimensional representation such as volume. This creates ambiguity of how could depth measurement be learnt with a meaning? How could infants perceive depth?

5.2.1 Depth Perception

One of the earliest explanations of depth perception is from Rene Descartes. Rene Descartes in year 1638 interpreted depth mathematically with relation to the binocular vision. Binocular vision builds a three-dimensional world with the slight perceptual difference between left and right eyes. Descartes described that such a visual difference forms parallel visual lines with far-away objects; contrarily, objects that are closer form acute angles of visual lines. Acuteness of visual lines perceived by both eyes could be used to determine how far away the object is.

Further findings have suggested that binocular vision develops between 3 to 5 months. Before this period, infants' right eye reacts from left to right while left eye reacts from right to left. Lange-Küttner (2018) concluded that binocular vision is developed after monocular vision. If the object was moving, infants could only track the object with monocular visions before 4 months old. If the object was not moving, infants could only react to static depth perception such as pattern and texture differentiation after 4 months. Hemker, Granrud, Yonas, and Kavšek (2010) have examined depth perception with static pictures in various textures and dimensions showing a difference in depth. They found that infants become sensitive to these static images with respect to depth differences between 5 and 7 months of age. These findings seem to agree that very early infancy could not exploit visible depth perception using binocular vision.

Berkeley (1910) raised a different proposal that insisted the engagement of tactile and motor experience to learn depth. Berkeley suggested that the size of an object shows a sign of how far the object is, the closer the bigger. With the help of the sensorimotor coordination, infants could raise their arms to try to reach the object. When the object approaches the subject, it finally touches the subject and could be perceived by haptic sense. The haptic sense signifies the closest distance between the object and the subject, i.e. zero depth. This suggestion has been examined with 5.5 months old infants that they are able to associate size with depth (Kotovsky & Baillargeon, 1998).

Discussion

Both arguments from Descartes and Berkeley do not disagree with each other. Descartes proposal implies that depth cannot be understood directly before 4 months old as binocular vision has not been formed yet. Given the fact that images projected onto retina is two-dimensional, depth measurement must be in some other form. Descartes' proposal supplemented this pitfall by measuring the acuteness of visual lines signals from both eyes. But this measurement does not come with a meaning automatically. Berkeley's suggestion can incorporate depth when it is zero with reference to the haptic sense. Depth changes are interpreted with size changes. Disregard of whether it is the size or the visual lines that matter, both interpretation does not reject that sensations provide some form of indirect measurements of "depth" as a way to identify it.

The question is how could such a change give rise to an understanding of depth changes. How can depth changes relate to a meaning in form of a unit that infants understand. The inputs of haptic sense can only be represented for the moment that the subject and object were in touch but not the changes of the size. Therefore, haptic sense cannot be the major inputs required for depth changes.

5.2.2 Eye-hand-mouth Coordination

Experiment 18 - Depth perception

One of the earliest responses to depth could be feeding. A few hours old infants seem to respond to feeding already. Futagi (2017) shows that young infants respond to an approaching finger by opening their mouth. Depth enables a response in an appropriate timing to react with very simple motion, i.e. mouth opening. However, food does not necessarily guarantee a response. From Rochat (1993), the sucrose delivery to the mouth of 8-weeks-old did not necessarily introduce a positive reaction to submit their hands towards their mouths. These seem to suggest that the haptic sense is inadequate to explain and interpret depth changes with a meaning.

Experiments of cliff detection intends to address the question of depth learning (Gibson & Walk, 1960). As young as 6 months old, very little infants would crawl across the illusionary cliff to respond to their mothers' calls. Schwartz, Campos, and Baisel Jr (1973) studied crawling infants and recorded their heart rate when they were exposed to shallow and deep sides of the illusionary visual cliff. They found that infants increased heart rate at 9-month-old when exposed to deep side of a cliff. It is a shift from 2 to 5-month-olds that had a decreased heart rate instead. They interpret this as a result of increased attention or interest that infants perceive depth but are fearless of the consequence. Authors concluded that young infants perceived depth but they do not understand its implication.

Discussion

These experiments show that infants do not understand depth even if infants can see depth. Although the understanding of depth might associate with various affectivity state such as danger (as a result of support learning for cliff events) and feeding, this must have to be learnt first.

They understand depth changes only when they can associate depth changes with their sensational changes. A crawl action induces a calibration of force exerted by the subject to move and change the depth perceived. This is when the subject can make sense out of depth as "farther away from me" or "close to me". Hence, the unit of depth changes can be compared to the unit of motion changes (such as force and speed) brought up by the subject. Once a calibrated conversion established, depth could be understood with reference to the subject.

This further echoes the finding of not understanding depth from previous chapter that monocular vision in early stage is only sufficient to keep track of motion involving horizontal and vertical components only. By keeping the object signals onto retina, the magnitude of the motion can actually be understood as the change of the eyeball movement caused by positional changes on retina but not depth. An approaching object was perceived with signals of occupying more and more of the retina sensory area only. Size changes are not a primitive measurement by a coordinate system and its corresponding meaning has to be learnt as well.

Experiment 19 – Depth Calibration

Further experiment has been conducted for 4-month-old that infants initially raised both of their arms to reach an object (Crichton & Lange-Kuttner, 1999) and then developed into one arm's reaching gradually afterwards. The experiment showed that they are not initially equipped with an accurate grasping and reaching behaviour. Reaching and grasping are the motor experiences that requires a learning of depth differences. How far do arms move? How much force do arms apply? How much strength could muscle provide? It seems to have a calibration period until they are certain about their own powerfulness.

Further evidence of calibration of the subject in respond to perceptional signals has been shown in the learning of walking. Babies just started walking has been found to be confused in relation to coordinate their motor actions. They might experience confusion for a period of time and result in recalibration in transition stage (Grzyb et al., 2012) when they acquired more capability to walk (Karasik et al., 2011).

Further depth calibration is found earlier before walking. Infants learn the appropriate angle to grasp an object after 5 months old (Lockman, Ashmead, & Bushnell, 1984). 5-month-old turned their hands into an appropriate orientation only when hands were close to the object while 9-month-old could prepare the orientation of their hands before stretching out their arms. This calibration process involves the anticipation of the location of an object relative to the subject (i.e. depth measurement and its conversion to motion), exploration of the travel distance of the hand, orientation of the target object and the way to properly exercise motor control. Thelen et al. (1993) explained that infants aged 3 to 5 months do not know how to calibrate and exercise their motor capability precisely initially. Infants failed to control their arms with appropriate energy and speed to reach the location of the target. This results in a reaching behaviour with both hands.

Early infancy cannot move their arms in a straight path towards the target to reach the object. This discontinuous trajectory motion of infants' arms during reaching also occurs when they cannot see their body and arms. Berthier et al. (2001) explained that reaching an object involves "eye movements, eye-head coordination, postural stability of the trunk, and eye-hand coordination"; it requires to look at the hand as well as the object for reaching initially. In older infants, they gradually developed into a smooth reaching action without fixing their vision at their hands. Therefore, through a calibrated movement of the subject, depth could be understood and reacted to.

Discussion

The interpretation of the Z values can be learnt through the change of how much force required to reach an object. This is a developmental process as infants' mobility is limited. Initially, infants are likely to manage visible depth most of the time with limited mobility as they cannot interpret measurable coordinates into a representation of actionable depth by their own motion. Upon the motor capability developed, coordination changes have to be re-interpreted and computed into attributes of a response such as force, speed and direction. Therefore, a calibration is required as the motor capability keep increasing. The calibration takes the output of a response back to input to re-adjust the scale of the conversion from perceivable change into subject's responsive magnitude. After that, the subject could understand the meaning of depth as a reference with the responses and subsequent sensor feedbacks.

The preference of using visual inputs as dominate inputs lead to a consequence of generation of cognitive questions to understand perceptual signals. If a visual perception has never been received for the form of sensory inputs received, a cognitive question can be generated and prioritised for actions. Action explores more properties of the object and receives more inputs to enrich the representation. For example, a call from mother stimulate the subject to turn over and crawl towards her. The tendency to engage action to complete visual representation could also be a form of drive to react and learn depth.

5.2.3 General Discussion

Gazing response cannot fully explain depth learning as the response only takes X and Y coordinate changes for outputs. Coordinate inputs are initially perceived in passive mode. A response to the perceived inputs can only be learnt with a mean through an understanding of the attribute changes initiated from the subject itself. An understandable mean could be the effort for a reach or a shift of the gaze. Motor actions such as reaching and crawling allow them to explore the relation of depth with certain amount of energy and force to get close to the object. The force being exercised could be linked with a comparison of the perceptual difference between their body and the object. By touching, they are able to exert active action onto the object to explore the consistency, such as hard and soft. The touches also give them a sense of pressure to signify a contact of the subject and an object. The meaning of depth could then be learnt through enrichment from haptic senses and the visual change of positions required from the subject to touch. A response required from the subject involves calibration period to convert the unit of the measurable coordinates into the force, energy and speed required for an action.

Multi-modal senses are dominated by visual preference. One may argue that a blind person could make use of their other sensations to "feel" the distance, for example, the strength of the smell or by means of haptic touches. Evidence has indicated that exploration of objects could be hindered without vision but can still be executed (de Bruin, Sacrey, Brown, Doan, & Whishaw, 2008). Therefore, the cognitive process that accommodate different senses as input should be able to accept inputs from various sense and produce a similar output to respond. However, if the coordinate system is initially used to create representation from visual perception, this implies that it should automatically create a cognitive question for visual inputs if no similar input can be found from memory for representation upon non-visual inputs.

5.3 Do Other Species Possess Similar Behaviours?

Other species were also found to be able to attach to their mother in a few exposures with a process of "imprinting". Lorenz (1937) referred imprinting as "an acquiring process" of "a personally recognised object" that develops social behaviours such as attachment. McFarland (2014) referred imprinting in general as "an aspect of learning that takes place during a sensitive period in the early stages of an animal's life". It characterises with a few exposures to an object in a specific period after birth and gives rise to following or attaching preference of animals. These strongly supported that imprinting is in itself a process of early learning and could be supported by new-borns with a straightforward process. Under what circumstances do other species form the attachment behaviour?

5.3.1 Early Stage of Learning

Chicken could form attachment with a familiar object even if it is not a chicken-like figure without physical contact. J. Wood (2013) incubated and raised chicks for 2 weeks in a dark and controlled chamber; virtual objects were projected on the walls of the chambers with different view point such as 60-degree smooth rotation. J. Wood (2013) tested the chicks with familiar and novel objects; Results showed that chicks stay with familiar objects longer. It seems to suggest that chicken could recognise the object and develop attachment behaviour if they were exposed to the object.

New-born chicks seem to have a learning process that is also largely sensitive to spatial arrangement initially characterised with smooth movement. Both J. N. Wood (2016) and J. N. Wood, Prasad, Goldman, and Wood (2016) raised that internal representation in chicks could be distorted due to rapid or non-smooth movement. The internal representation with reference to the imprinting behaviour of new-born chicks was further investigated in J. Wood and Wood (2018); they suggested that chicks could identify objects and tolerate angle changes if the object movement was smooth; if objects were perceived with non-smooth movement, chicks tended to stay with the familiar views and become less sensitive to stay with familiar object. J. Wood and Wood (2018) conducted a similar test as in J. Wood (2013); the projected object was rotating either smoothly or non-smoothly (i.e. different views of the object are projected randomly instead of in rotation sequence) with 30 frames per second or 1 frame per second in initial phase and then chicks were tested against different viewpoint of rotated novel objects. It is found that smooth rotation enables chicks to recognise and stay with familiar object; the time spent with familiar object by chicks increased over time with smooth rotation if testing is repeated but not non-smooth one. Therefore, J. Wood and Wood (2018) concludes that chicks could perceive and represent familiar objects with a smooth motion. In some further experiments, chicks are not just capable of recognizing familiar objects by spatial configuration but also seems to be able to bind color into the internal representation. J. N. Wood (2014) exploited the same procedure to imprint incubated chicks with colored shapes; chicks are then further tested against objects in familiar and novel colours. Findings suggested that these chicks could differentiate colour differences. Even with chicks, featural configuration has a similar role in visual learning as with human infants.

This strongly suggested that visual perception has significant influence in imprinting to form attachment behaviour. Harlow (1961) pinpoints that this learning does not initiate with a particular innate choice but could be directed to any objects. Harlow (1960) conducted experiment with monkeys; monkeys were tested with two mothers – one with food and one does not – and they shared approximately equal amount of time with either one. Therefore, the process does not seem to be reinforced by food (Hinde, 1961). Neither nursing is essential to form attachment (Salzen, 1962). These suggest imprinting is more depended on exposure.

While some researchers argued that imprinting cannot be changed, R. A. Hinde, W. H. Thorpe, and M. A. Vince (1956) provide evidences of imprinting in birds that they might follow multiple different models as well as changeable attaching behaviour. Rosenblatt, Turkewitz, and Schneirla (1961) investigate kittens for its attachment strength; they took kitten away at various time from mother and isolated them with an auto-feeder. Rosenblatt et al. (1961) found that this isolation prevents kitten from suckling mothers again after the isolation, especially those in younger age. This evidence echoes the finding from R. A. Hinde et al. (1956) that imprinting characterised with first exposure can be interfered later on by different circumstances.

Discussion

Imprinting as a result of perceptual learning characterized by motion-driven perceptual processing and an expressive form of attachment to the object. Simpler species such as chicken can also learn to represent objects with partial properties to form attachment. Salzen (1978) argued that the similarity between imprinting and attachment are: i) building a close relationship with the target; ii) satisfying a maximization of positive affectivity and reduction of negative affectivity; iii) exhibiting many interactions between the target and the subject. The intimate relationship enables more interaction. These interactions can then allow a construction of a rich representation of the attached target. A rich representation allows early infancy and simpler species to recognize attachment target easily for other purposes.

From these experiments, the rich representation required by imprinting and attachment are both supported by a coordinate system to process spatial changes. If this is the case, what differ human beings from other species in terms of cognitive process?

5.3.2 Cognition Development in Other Species

Regarding more advance cognitive behaviour, dogs and wolves demonstrate certain degree of sensorimotor intelligence and object permanence as in human beings. They react to stimulation and action against it with some degrees of representation comparison. Horowitz (2014, pp. 155-174) compared the cognition process between wolf and dog using a framework adopted from Piaget (1954); the research investigated cognitive development of dogs and wolves and their stage of sensorimotor intelligence and degree of object permanence. Four grey wolves aged 10 days were i) observed for their interactions with different objects (e.g. puppets, towels, cardboard boxes and ropes), and ii) tested to seek a hidden object from three boxes in different colours. It is found that wolves cannot develop their sensorimotor intelligence beyond stage 4 as exhibited in Table 4 stages of sensorimotor intelligence, i.e. coordinate two different actions to solve a problem. They failed to demonstrate a trial and error exploratory reactions to find interesting consequences; compared with Piaget's observations, human beings excel in coordinating the sensorimotor behaviours even without direct perceptual signals received.

Stage 1	Reflex behaviours
Stage 2	Habits were adopted from repeated reflexes such as walking, scratching.
Stage 3	Habits from repeated actions induce influence to surrounding environment such as tugging a log.
Stage 4	Goal seeking with combination of two actions such as biting a toy or a fellow are exhibited.
Stage 5	Trial and error exploratory actions that discover external effects.
Stage 6	Without direct stimulation, series of actions are applied to explore their effects and/or achieve a goal.

Table 4 stages of sensorimotor intelligence

From the perspective of invisible objects processing, dogs and wolves aged 11 weeks could successfully seek the hidden objects from three boxes in triple visible displacement test, i.e. stage 5b as shown in Table 5. Wolves failed to achieve the single invisible displacement test, i.e. stage 6a. The test hid a target object into a transportation device which was then shuffled amongst three boxes. The object inside the device was invisibly transferred to one of the boxes. Then experimenter showed wolves with the empty transportation device before placing it into another box without the target object. Wolves failed to seek the object that was displaced invisibly from the device to the box.

Stage 3	An object was partially hidden behind a box.
Stage 4a	An object was placed behind the same target box visibly each time after a wolf was released.
Stage 4b	An object was placed behind the same target box visibly each time when a wolf was released.
Stage 5a	An object was placed behind a box that is different from previous stage.
Stage 5b	An object was firstly hidden behind a box then was visibly transferred to another box. Then, the object was further visibly transferred to the third box. Each box was exploited for hiding the object once.
Stage 6a	An object was visibly placed inside a transportation device. The device was then moved behind one of the boxes. Experimenter invisibly transferred the object to the box and showed wolf the empty device. Then, the device was further moved behind another box. The object was not hidden in the same box as in stage 3 and 4.

Table 5 stages of object permanence

5.3.3 General Discussion

Imprinting seems to be able to explain very early learning with simpler mechanism and a straightforward process. First exposure is deterministic of what to cling to. It could be a consequence of richest representation construction available for a particular object that has been updated with many cognitive responses by proximal interactions. This is similar to human beings in early infancy that they construct representation and form representation out of familiar objects. There is evidence that this form of attachment happens even when there is no animated object (Harlow, 1960). Birds not only showing imprinting behaviour but also good imitation capability which is believed to be the result of their superior visual acuity when compared to other species (Zentall, 2004). Hence, imprinting as first capture of a novel object to be familiarised with, might be largely related to the richness of a representational construction in first sight. The first exposure can subsequently be changed but takes more exposures. Exposures allow representation system to enable more partial recognition and construct a richer representation based on previous representation. The least uncertain object is always the attached target with least effort of recognition. It could be the criteria as a result of obtaining positive affectivity, e.g. sense of security.

However, these evidences have put up a question of what make human beings differ from simpler species. More evidences suggested that human beings excelled in responses involving more levels of representation and relationship constructions when no direct input is given. Human beings can process and respond when the perceptual signal discontinued. Human beings can also infer circumstances without direct perception at all. The characteristics of cognitive process without direct stimulation could be the capability of building a representation from previously perceived inputs for more representational comparison of spatial relations and resolution of cognitive questions. To be able to resolve more advance cognitive problems out of invisible physical phenomenon, it pre-requires the capability to raise more questions and construct representations and its relationships out of lower level of representations and their relationships. For example, being able to represent the invisible object placed in the transfer device, update the positional change of the object placed inside when it is shown as empty device and seek the object with a representation from a representation along possible motion path carried by the transfer device. Simpler species do not seem to be able to build representation out of another representation steadily.

The processes underneath the same visual learning behaviours between chicken, canines and human beings might start with a similar mechanism but differ with more advanced representational computation. Human beings are more sophisticated in the processes of filling in the gap from perception, raising cognitive questions and resolve the questions from representation and relations from previously knowledge. Human beings do not necessarily rely on perceivable inputs. This differentiation promotes advanced learning from primitive early mechanism.

6 Hypothesis of Initial Learning Model

Summarizing the empirical evidences of occlusion events, an object could be learnt with the following characteristics: i) an object could be tracked through its motion; ii) infants could recognise and respond to stimulation exploiting whatever properties being perceived; iii) responses could be computed with a measurement understandable to the subject. Two major hypotheses can be drawn from Chapter 3 and 4 for physical object learning:

i) A representation of an object must exist in order to compare a moving object with an occluder;

ii) A response must convert representation of perceivable objects and calibrate properly through the measurement in form of a unit exercised by the subject initially.

The following model attempts to address early visual learning with representation formation in order to investigate if it could replicate the behaviours exhibited in empirical experiments and also prove the hypotheses listed above.

6.1 A Preliminary Model

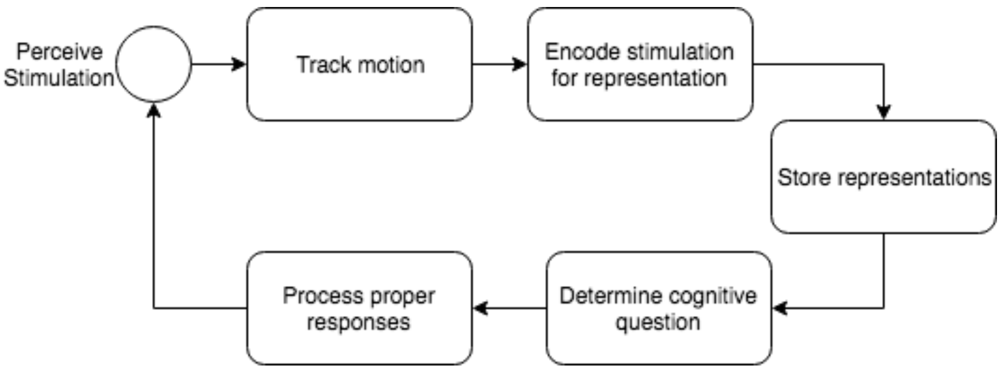

Figure 24 Process of Visual Learning

This hypothetical model addresses visual learning process by early infancy with a simple process as shown in Figure 24. The process focuses in four areas: i) tracking process; ii) storage structure; iii) inputs for representation; iv) inputs for recognition; v) outputs for responses.

6.1.1 Tracking Process

From empirical evidences, the characteristics of tracking are: i) predictive; ii) track with representational object for comparison enablement. Tracking of an object works with snapshots. To catch the image, the subject responds by shifting the gaze or turning the head. The subject featured with that eye-scanning is not sensitive to everything perceived but approximately chasing after the attentive parts of the object (Johnson, Slemmer, & Amso, 2004; Rucci & Victor, 2015). The eye-scanning or gaze responses from infants does not form a precise trajectory. Hence, it is wrong to provide a function to predict a path accurately. This tracking strategy takes a simple process to respond to next location of the trajectory and adjust according to the discrepancy. The smoothness of the motion of a group of features against time could then be understood as a regular change of time interval that could be predicted from last positional change. This could contribute to better learning, as demonstrated in chicken (J. Wood & Wood, 2018).

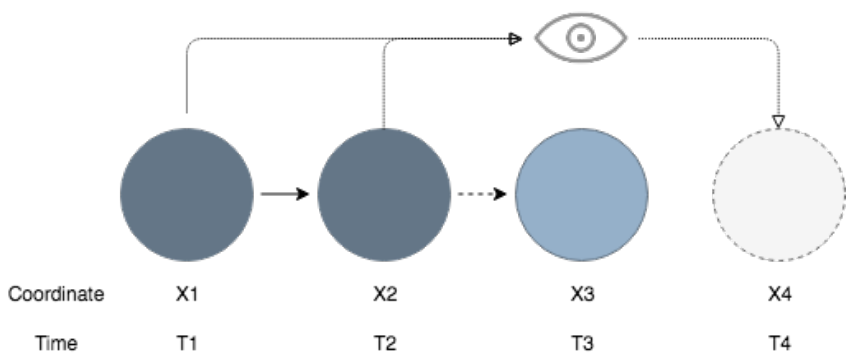

Figure 25 Gaze shifting mechanism

As illustrated in Figure 25, if object moved from X1 to X2 then X3 at time T1, T2 then T3, infants were gazing X3 at T2 while the object was moving from X2/T2 to X3/T3 and computing X4 with the magnitude change of position and direction from X1 to X2 given the time change from T1 to T2. When object arrived at X3, they have already computed X4 and shifted their gaze. Therefore, infants' gaze shifting behaviours take internal representation of previous snapshot at a given time recently to compare with current position and move ahead. Without a previous reference, infants can only chase after the moving object when it is already moved. Otherwise, they could only respond to a predicted motion from certain reference points and might lose the track. The memory of representing previous snapshots and their changes provides previous representations for prospective gaze shifting.

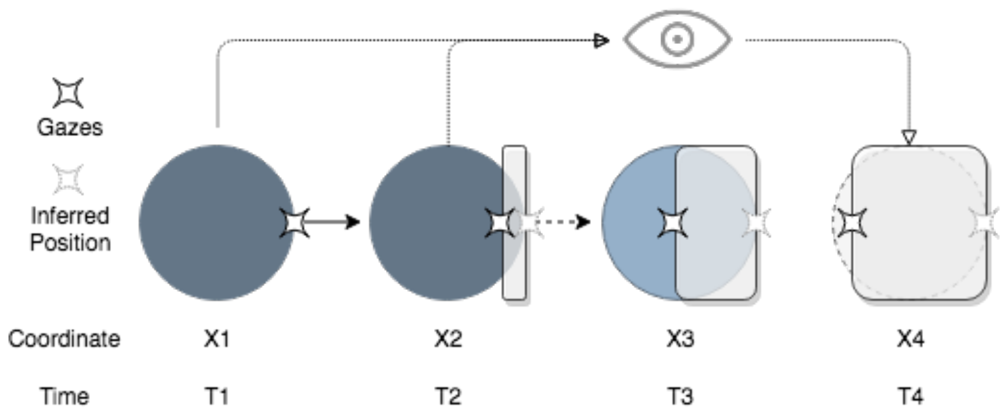

Figure 26 Occlusion and Gaze

If the occluder was narrow, the object would be tracked when it was invisible. As illustrated in Figure 26, if part or most part of an object was being occluded, the highly contrasted feature would be between the connection around the edge of the occluder and the moving object. The original point of reference have to be computed from what could be perceived by a comparison to the represented object. By that, the current position, which is perceived with the properties of an occluder as inputs, could be computed and gazed at to keep tracking.

6.1.2 Storage Structure

Working with the processing capability is the memory. Humans possess declarative and non-declarative memories that manage various activities (Ross, 2015); with the explicit declarative memory, further divided into short term and long term as indicated. Short term memory serves as a working memory that carries information being received in a few seconds and works as a push-down stack so that the most remote one might be discarded. It is found that the working memory has specific stores for verbal and spatial information and each piece of "data" is stored separately. Long term memories are managing events and facts that gradually sublime into "facts". When a patterned "fact" matches the memory, the signal transmission would be stronger and more reliable, hence, enhancing the relay and response for a robust recall of memory. Although what does it mean by "fact" could not be defined in this research, it is clear that recent and remote memory are separate.

Brain is believed to pull together different independent "facts" by a temporary recall of memories of the medial temporal lobe; research states that the medial temporal lobe would not be participating in very remote memories after learning and reorganization of the facts (Squire & Zola-Morgan, 1991). Studies on amnesia patients suggest that different parts of the brain work in parallel for declarative and non-declarative memories. Squire and Zola (1996) aggregated the lesion studies of amnesia patients and suggested that knowledge influenced by previous acquaintances such as priming was still functional even if declarative memory was impaired. Our brain has an independent department for each of the functional use.

The approach of loosely linking properties groups should be taken such that each of them could be recognized on its own. Relations of episodes and other properties are required to be stored. Structured facts or organized hypothetical change could form a rule in separate storage and become the type of more salient memory that could facilitate memory retrieval and consolidation to resolve problems (Weingartner, Sitaram, & Gillin, 1979). From these neuroscience studies, humans demonstrate the ability to access the rule and the recent remembered snapshots separately. Therefore, snapshots information may have a link to the object but could also be a separate representation. The storages are stored separately and hierarchically as: i) groups of properties; ii) snapshots (episodes); iii) rules.

6.1.3 Inputs for Representation

Amongst all of the evidences, positions of perceivable properties have a significant role to construct an object. Infants even two hours old could already recognise shapes (Leslie B. Cohen & Younger, 1984). The perceivable signals of these shapes do not necessarily come as connected form. An understanding of an object starts with a computation of coordinates of those properties that moved together in the same motion. When an object projects onto the retina, the process could articulate the signals but is unable to describe its meaning yet. Signals themselves do not explain what a movement is but the same positional differences of the properties can be defined in form of a measurement of the position shifts on the retina.

One important evidence of defining an object being tracked is that not all the details of the object are being remembered. Instead of understanding all aspects of the object to enable the process, an object is only briefly recognized by spatial information with or without feature details. The most obvious characteristics being that is shape. The rough coordinates of shapes are remembered as in the mooney face experiment (Leo & Simion, 2009b). It could be interpreted as a composition of some properties linked together with their positions. The positional differences of each properties are the basic relation to construct various representations.

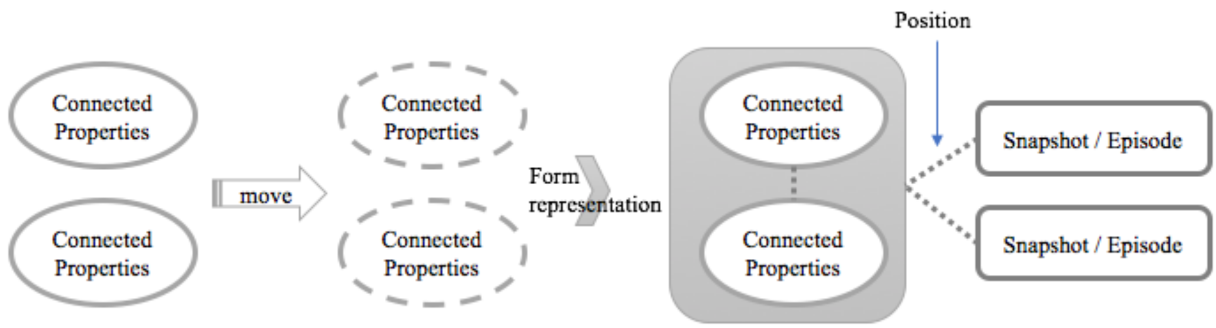

Figure 27 Initial formation of a representation of an object

Construction of an object initially groups connected properties with a relation of similar positional change (Figure 27). Their positions are used to relate different connected properties groups as one object. Subsequently, the object is related back to the snapshot of the episode with its position.

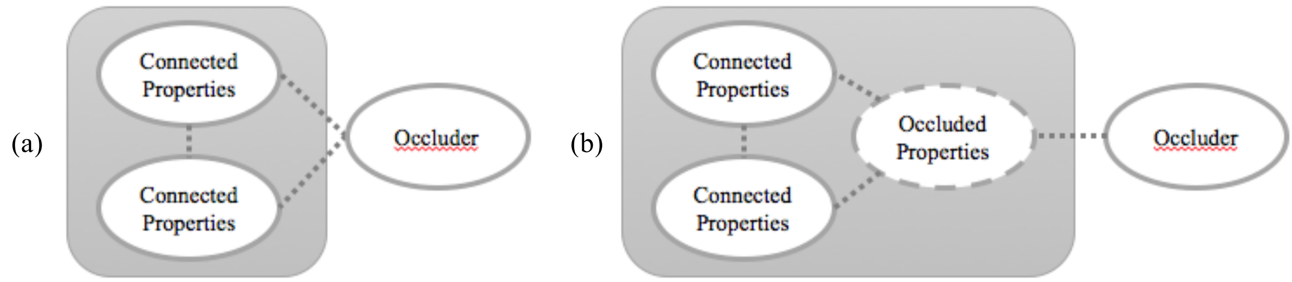

Figure 28 Representation of an object with occluded properties. (a) A representation without formation of a question about what are the occluded properties; (b) A representation with a doubt on what is the occluded properties.

Representation for very early stage only processes visible inputs if the occluder is too wide to track. The representation is composed of what is being perceived. Perceived properties moved together are grouped as one object while a relation of positional difference built with the stationary occluder (Figure 28 (a)). In the case of a representation of hidden portion of the object required, a representation is required by a tendency to connect properties if there is a positional gap. The representation formed for the object created a questionable properties awaiting update (b). The questionable properties are linked with the occluder with a positional relation. Therefore, when an input for the object was required, the representation of the occluder could be used to complete the object representation. Subsequently, this forms the basic representation of an object that enables tracking in occlusion events.

6.1.4 Inputs for Recognition

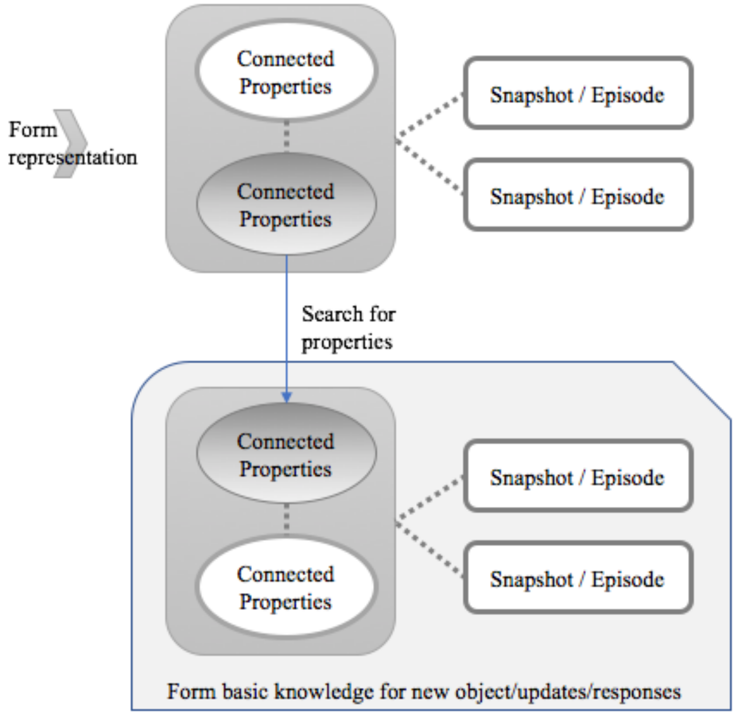

Figure 29 Representation formed from perception will be searched through memory for previous representations

When an object was formed with a representation, its properties group could be represented to search for previous knowledge related (Figure 29). In case of discrepancy of other properties, a question could be raised for responses to explore. More exposure gains more perceptions and certainty of the properties. A particular group of properties perceived and exploited more frequently that could result in a stronger identifier. The stronger the properties group, the higher the priority is for recognition. This forms high salient properties for recognition.

Recognition also takes a more recent memory of snapshot to make a comparison. In other word, Strength of the property could be reduced against time elapsed. The repetition of a matching properties group forms a stronger strength to attach to the object. A difference with previous property could be described as a weak property and could provide a branch of the object properties. Hence, this could utilise a group of stronger properties of the object for recognition and allows discrepancy of other fine details. Objects that are with richer and stronger representation would be most likely used for recognition, hence, shape alone might not be efficient to identify a face considering it could be turning around.

6.1.5 Outputs for Responses

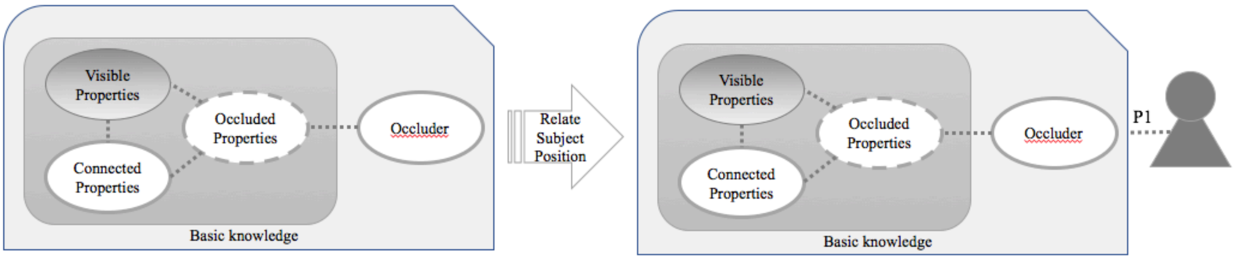

Figure 30 Perceived properties has to be related back to the subject

Before a response could be reacted disregard of shifting eyes or reaching by hands, the perceived measurements have to be understood by the subject for a response. First step is to relate the positional differences with the subject as P1 in Figure 30. This could then be used to compute the positional relationship of the object.

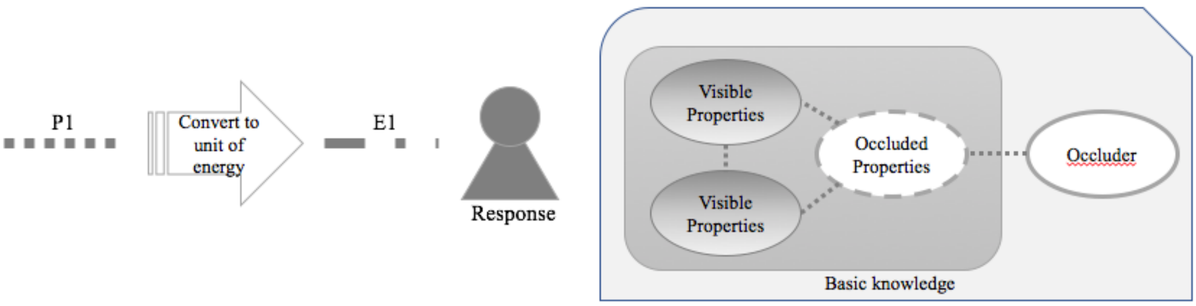

Figure 31 The position between the object and the subject (P1) has to convert to the unit of energy required for a response

Subsequently P1 has to be converted to the unit of energy required for an action, e.g. force required to reach the hands or the number of shifts required by eyes. Upon reception of update of perceptual signals, more properties groups could be matched to verify if this is the desirable object to be used or create a new representation (Figure 31). The tendency to connect properties moved together groups the disconnected properties into one object. Hence, if the object reappeared from occlusion as two objects (e.g. two rods), the grouping of the object had to be restructured with the occluded properties removed. If it was a complete rod, the properties groups were simply merged together by removal of the positional relations. Outstanding properties as a questionable form stirred up an "interest" for responses, for example, the occluded properties. If it was a strong properties group, an action could be executed to find the answer, for example, responding to the call from a mother.

A process could then be constructed with minimal mechanism to chain the tracking process to perceive properties for representational construction, storing the representations, recognizing more objects, figuring out outstanding cognitive questions and supplying outputs for responses in a closed loop.

7 Conclusion and Future Works

The preliminary model for object representations based on the rich empirical evidences of occlusion examines the process of how could perceivable images be perceived and identified as an object. A learning model of early infancy does not just gain knowledge, it also copes with the highly complex immediate environment flexibly and structurally. It could accept different types of inputs from sensors to extend the description of object knowledges. The importance of being able to structurally support new inputs in relation to the subject is to be able to accommodate physical growth that adapts changes and leads to more perceptual learning.

The process that works with execution of capturing inputs and generation of outputs can accommodate various complexity of inputs and outputs. One of such execution is caused by physical growth. Physical change induces calibration progressively and reacts with a different output. Mere execution does not give rise to a meaning or understanding of what has been perceived. The process enabling the learning has to be constructed and reacted with various degree of maturity of inputs. This is not just an increase of executional power but also exhibiting an adaptation for knowledge enrichment. Consequently, the growth of intelligence could then be gained progressively and responsively by increasing the power of input capture against physical stimulations with the assumption of basic capability in Chapter 4.1.

Infants' inferential capability is not only about prediction accuracy but also embedded with a process to figure out cognitive questions with representations. This suggested that the function of cognitive process is not just about responding with an action to a goal generated from other process but also raising relevant questions as a goal. The questioning capability aims to explore further sensory signals to ensure continuous supplies of inputs and provide adequate instruction for responses.

Other animals have failed to advance into higher level responses without direct stimulation. This indicates the capability to build representation out of representation without direct perception is most critical to advanced intelligence. Being able to construct an internal representation properly could be just the first step of the process. This model serves as a starting point to explore higher level cognitive behavior that is developed with insufficient inputs like language learning as the model does not required any mandatory properties from sensors but coordinates only. Internal representation construction forms a complicated network of knowledge by loosely coupled relations. The relation amongst different sensory signals could induce a conversion process and might build interesting implication to refer to various representation for behaviors that seem to apply advanced rule.

This research does not intend to infer all common-sense formations with the same mechanism. I would like to point out that the ignorance of a proper object representation had created barrier to develop advance cognitive behavior under the light of how dynamic physical objects could be and how species grow organically with changing environment.

Limitation

This research explores possible input and desirable output to construct a simple process using Yeap's approach. Due to time limitation, the model might not be a full coverage of what and how species could learn visually. Further works of detail process, algorithm and implementation are required. The hypotheses should be verified by comparing the result with psychological experiments. Verification of the model has to demonstrate two behaviours: i) behaviours as exhibited in similar empirical experiments with human infants; ii) emergence of novel object representation. If the response mimics similar findings as psychologists measured and able to continue learning new objects, the model could shed lights to cognitive development of visual learning in early infancy. Through an implementation and replication of the empirical experiments, it could be validated against how close could the model explain visual learning of early infants.

Future Works

In terms of visual learning, there are other forms of physical phenomenon to be explored such as orientation, solidity and further exploration of containment events. Similarly, just like occlusion, they could be a developmental process that spanned across different phases and required careful examination of what could be perceived initially and how it grows into various representations.

Many works focused on one kind of perceptual signals. To understand behaviors exhibited by intelligent species, study of inputs is necessary to construct input-representation-output relationship. Building the representations and their relations required careful studies of what could be perceived and responded during very early infancy. No matter which approach researchers adopted, most work focused on specific sensation but very few works addressed multi-modal effect. Each sensation does have its own advantage and disadvantage. Focused on one type of sensational signals would prevent us to address possible learnings that could be learnt through other senses and strengthened through representational relations of multi-modal effect, for example, haptic sense.

Haptic sense might be able to explain learning of some physical rules that could be hard to explain by visual learning. Being able to touch an object could be an important part of initial infant stage such as grasping and eating. By means of touching, infants would be able to feel the principles of force in measurable units to form haptic inputs, e.g. pain and pressed. The advantage of haptic inputs is its highly abundance sensors that was designed to enable the sense of touch only when an object is closed enough. Haptic sense does not tell us anything about positional changes. Given many evidences pointed out that spatial information is very significant to early stage of learning. Haptic sense seems to be useless from this point of view. If the other sensorimotor helped to pull or push things toward skins, haptic information came at last with respect to distance. Then, a contact with reference to something the subject understood (touch) happened. A contact can then be represented and measured by the haptic sense. Being too focused in visual signals might be the barrier to represent the meaning of a contact and might result with different behaviors. Some rules might have to be derived from multi-sensory inputs.

Rule construction and its representation has not been investigated in this research. Common sense rules could be applied to solve cognitive question and respond to stimulation in a top-down manner. Emergence of rules might differentiate intelligence showcased in different species. It could explain why certain species behave and respond as they are in later stage. Rule formation is critical to the emergence of advanced cognitive behaviors.

Cognitive process features with bottom-up learnings and top-down application of rules that forms a closed loop to react to changing environment. The key principle of the process is to address “change”. Vision detects critical changes and becomes major guidance for body to react due to its accurate coordination inputs. If earliest response is the ability to look at things, species need a coordinate system minimally to locate things to continue looking at things. Therefore, the coordinates system could be the earliest system in visual learning.

8 Reference

Aguiar, A., & Baillargeon, R. (1999). 2.5-Month-Old Infants' Reasoning about When Objects Should and Should Not Be Occluded. *Cognitive Psychology, 39*(2), 116-157. https://doi.org/10.1006/cogp.1999.0717

Aguilar, W., & Perez y Perez, R. (2015). Dev E-R: A computational model of early cognitive development as a creative process. *Cognitive Systems Research, 33*, 17-41. https://doi.org/10.1016/j.cogsys.2014.09.002

Alvesson, M., & Sandberg, J. (2013). *Constructing research questions : doing interesting research*: London : SAGE, 2013.

Bahrick, L. E., Lickliter, R., & Castellanos, I. (2013). The development of face perception in infancy: Intersensory interference and unimodal visual facilitation. *Developmental Psychology, 49*(10), 1919.

Bailey, T., & Durrant-Whyte, H. (2006). Simultaneous localization and mapping (SLAM): part II. *IEEE Robotics & Automation Magazine, 13*(3), 108-117. https://doi.org/10.1109/MRA.2006.1678144

Baillargeon, R., & DeVos, J. (1991). Object permanence in young infants: Further evidence. *Child Development, 62*(6), 1227-1246.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20*(3), 191-208. https://doi.org/https://doi.org/10.1016/0010-0277(85)90008-3

Bennett, A. T. (1996). Do animals have cognitive maps? *Journal of Experimental Biology, 199*(1), 219-224.

Berkeley, G. (1910). A new theory of vision, 1709. *Everyman, 483*(117), 53.

Bertenthal, B. I., Gredeback, G., & Boyer, T. W. (2013). Differential Contributions of Development and Learning to Infants' Knowledge of Object Continuity and Discontinuity. *Child Development, 84*(2), 413-421. https://doi.org/10.1111/cdev.12005

Bertenthal, B. I., Longo, M. R., & Kenny, S. (2007). Phenomenal Permanence and the Development of Predictive Tracking in Infancy. *Child Development*(1), 350.

Berthier, N. E., Bertenthal, B. I., Seaks, J. D., Sylvia, M. R., Johnson, R. L., & Clifton, R. K. (2001). Using Object Knowledge in Visual Tracking and Reaching. *Infancy, 2*(2), 257-284. https://doi.org/10.1207/S15327078IN0202_9

Berthouze, L., Prince, C., Littman, M., Kozima, H., & Balkenius, C. (2007). Constructivist Anticipatory Learning Mechanism (CALM)–dealing with partially deterministic and partially observable environments.

Bowlby, J. (1958). The nature of the child's tie to his mother. *The International journal of psycho-analysis, 39*, 350.

Bowlby, J. (1969). Attachment, Vol. 1 of Attachment and loss: New York: Basic Books.

Bremner, J. G., Johnson, S. P., Slater, A., Mason, U., Foster, K., Cheshire, A., & Spring, J. (2005). Conditions for young infants' perception of object trajectories. *Child Development, 76*(5), 1029-1043.

Bremner, J. G., Mason, U., Cheshire, A., Spring, J., Johnson, S. P., & Slater, A. (2007). Conditions for young infants' failure to perceive trajectory continuity: PAPER. *Developmental Science, 10*(5), 613-624. https://doi.org/10.1111/j.1467-7687.2007.00616.x

Bremner, J. G., Slater, A. M., & Johnson, S. P. (2015). Perception of Object Persistence: The Origins of Object Permanence in Infancy. *Child Development Perspectives, 9*(1), 7.

Bronson, G. W. (1990). Changes in infants' visual scanning across the 2- to 14-week age period. *Journal of Experimental Child Psychology, 49*(1), 101-125. https://doi.org/https://doi.org/10.1016/0022-0965(90)90051-9

Bulf, H., & Turati, C. (2010). The role of rigid motion in newborns' face recognition. *Visual Cognition, 18*(4), 504-512.

Bushneil, I. W. R., Sai, F., & Mullin, J. T. (1989). Neonatal recognition of the mother's face. *British Journal of Developmental Psychology, 7*(1), 3-15.

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development, 10*(1-2), 67-74.

Bushnell, I. W. R. (2003). New born recognition. In O. Pascalis & A. Slater (Eds.), *The Development of Face Processing in Infancy and Early Childhood: Current Perspectives* (pp. 41-53). New York: Nova Science Publishers, Inc.

Cernoch, J. M., & Porter, R. H. (1985). Recognition of maternal axillary odors by infants. *Child Development*, 1593-1598.

Chaput, H. H. (2004). *The constructivist learning architecture: A model of cognitive development for robust autonomous robots*.

Chomsky, N. (1979). *On language : Chomsky's classic works language and responsibility and reflections on language* [Electronic document]: New York, New York ; London, [England] : The New Press, 1979.

Cohen, L. B., Chaput, H. H., & Cashon, C. H. (2002). A constructivist model of infant cognition. *Cognitive Development, 17*(3-4), 1323-1343. https://doi.org/10.1016/S0885-2014(02)00124-7

Cohen, L. B., & Younger, B. A. (1984). Infant perception of angular relations. *Infant Behavior and Development, 7*(1), 37-47.

Crichton, M. T., & Lange-Kuttner, C. (1999). Animacy and propulsion in infancy: tracking, waving and reaching to self-propelled and induced moving objects. *Developmental Science, 2*(3), 318.

Csibra, G. (2001). Illusory contour figures are perceived as occluding surfaces by 8-month-old infants. *Developmental Science, 4*(4).

de Bruin, N., Sacrey, L.-A. R., Brown, L. A., Doan, J., & Whishaw, I. Q. (2008). Visual Guidance for Hand Advance but Not Hand Withdrawal in a Reach-to-Eat Task in Adult Humans: Reaching Is a Composite Movement. *Journal of Motor Behavior, 40*(4), 337-346.

Ding, Y.-h., Xu, X., Wang, Z.-y., Li, H.-r., & Wang, W.-p. (2012). Study of mother–infant attachment patterns and influence factors in Shanghai. *Early Human Development, 88*, 295-300. https://doi.org/10.1016/j.earlhumdev.2011.08.023

Drescher, G. L. (1989). *Made-up minds: a constructivist approach to artificial intelligence*. Massachusetts Institute of Technology.

Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping: part I. *IEEE Robotics & Automation Magazine, 13*(2), 99-110.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Physics of life reviews, 11*(3), 329-364.

Franz, A., & Triesch, J. (2010). A unified computational model of the development of object unity, object permanence, and occluded object trajectory perception. *Infant Behavior and Development, 33*(4), 635-653. https://doi.org/10.1016/j.infbeh.2010.07.018

Freudigman, K. A., & Thoman, E. B. (1993). Infant sleep during the first postnatal day: an opportunity for assessment of vulnerability. *Pediatrics, 92*(3), 373-379.

Futagi, Y. (2017). Eye-Hand-Mouth Coordination in the Human Newborn. *Pediatric Neurology, 75*, 43-47. https://doi.org/10.1016/j.pediatrneurol.2017.06.009

Gava, L., Valenza, E., Turati, C., & Schonen, S. d. (2008). Effect of partial occlusion on newborns' face preference and recognition. *Developmental Science, 11*(4), 563-574.

Gibson, E. J., & Walk, R. D. (1960). The visual cliff. *Scientific American, 202*(4), 64-71.

Graves, A., Wayne, G., & Danihelka, I. (2014). Neural turing machines. *arXiv preprint arXiv:1410.5401*.

Gronqvist, H., Gredeback, G., & Hofsten, C. V. (2006). Developmental asymmetries between horizontal and vertical tracking. *Vision Research, 46*, 1754-1761. https://doi.org/10.1016/j.visres.2005.11.007

Grzyb, B. J., Del Pobil, A. P., & Smith, L. B. (2012). How walking influences the development of absolute distance perception (Vol. 7426 LNAI, pp. 65-74).

Guerin, F. (2011). Learning like a baby: a survey of artificial intelligence approaches. *The Knowledge Engineering Review, 26*(2), 209-236. https://doi.org/10.1017/s0269888911000038

Harlow, H. F. (1960). Affectional behavior in the infant monkey. *Central nervous system and behavior*, 307-357.

Harlow, H. F. (1961). The development of affectional patterns in infant monkeys.

Hayashi, M., & Matsuzawa, T. (2017). Full length article: Mother–infant interactions in captive and wild chimpanzees. *Infant Behavior and Development, 48*(Part A), 20-29. https://doi.org/10.1016/j.infbeh.2016.11.008

Hemker, L., Granrud, C. E., Yonas, A., & Kavšek, M. (2010). Infant perception of surface texture and relative height as distance information: A preferential-reaching study. *Infancy, 15*(1), 6-27.

Hendriks, A. W., van Rijswijk, M., & Omtzigt, D. (2011). Holding-side influences on infant's view of mother's face. *Laterality, 16*(6), 641-655.

Hespos, S. J., & Baillargeon, R. (2001a). Infants' Knowledge about Occlusion and Containment Events: A Surprising Discrepancy. *Psychological Science, 12*(2), 141-147.

Hespos, S. J., & Baillargeon, R. (2001b). Reasoning about containment events in very young infants. *Cognition, 78*(3), 207-245. https://doi.org/https://doi.org/10.1016/S0010-0277(00)00118-9

Hespos, S. J., & Baillargeon, R. (2008). Young infants' actions reveal their developing knowledge of support variables: Converging evidence for violation-of-expectation findings. *Cognition, 107*(1), 304-316. (Original work published Cognition)

Hinde, R. A. (1961). The establishment of the parent-offspring relation in birds, with some mammalian analogies. *Current problems in animal behaviour*, 175-193.

Horowitz, A. (2014). *Domestic Dog Cognition and Behavior : The Scientific Study of Canis Familiaris* [Book]. Heidelberg: Springer.

Hossain, M., & Yeap, W. K. (2013). How Albot1 Computes its Topological-metric Map. *Procedia - Social and Behavioral Sciences, 97*, 553-560. https://doi.org/https://doi.org/10.1016/j.sbspro.2013.10.272

James, W. (1890). The Principles of Psychology. *2*.

Johnson, S. P., & Aslin, R. N. (1996). Perception of object unity in young infants: The roles of motion, depth, and orientation. *Cognitive Development, 11*(2), 161-180.

Johnson, S. P., Bremner, J. G., Slater, A., Mason, U., Foster, K., & Cheshire, A. (2003). Infants' Perception of Object Trajectories. *Child Development*(1), 94.

Johnson, S. P., Bremner, J. G., Slater, A. M., & Mason, U. C. (2000). The Role of Good Form in Young Infants' Perception of Partly Occluded Objects. *Journal of Experimental Child Psychology, 76*(1), 1-25. https://doi.org/https://doi.org/10.1006/jecp.2000.2545

Johnson, S. P., & Nanez, J. E. (1995). Young infant's perception of object unity in two-dimensional displays. *Infant Behavior and Development, 18*(2), 133-143. https://doi.org/https://doi.org/10.1016/0163-6383(95)90043-8

Johnson, S. P., Slemmer, J. A., & Amso, D. (2004). Where infants look determines how they see: Eye movements and object perception performance in 3-month-olds. *Infancy, 6*(2), 185-201. https://doi.org/10.1207/s15327078in0602_3

Kangasniemi, M., Utriainen, K., Ahonen, S.-M., PietilÃ¤, A.-M., JÃ¤Ã¤skelÃ¤inen, P., & Liikanen, E. (2013). Narrative literature review: from a research question to structured knowledge. *Hoitotiede, 25*(4), 291-301.

Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition From Crawling to Walking and Infants' Actions With Objects and People. *Child Development, 82*(4), 1199-1209. https://doi.org/10.1111/j.1467-8624.2011.01595.x

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*(4), 483-524. https://doi.org/https://doi.org/10.1016/0010-0285(83)90017-8

Kelly, D. J., Liu, S., Lee, K., Quinn, P. C., Pascalis, O., Slater, A. M., & Ge, L. (2009). Development of the other-race effect during infancy: Evidence toward universality? *Journal of Experimental Child Psychology, 104*(1), 105-114.

Kotovsky, L. (1994). 2.5-month-old infants' reasoning about collisions

Kotovsky, L., & Baillargeon, R. (1994). Calibration-based reasoning about collision events in 11-month-old infants. *Cognition, 51*(2), 107-129.

Kotovsky, L., & Baillargeon, R. (1998). The development of calibration-based reasoning about collision events in young infants. *Cognition, 67*(3), 311-351.

Lange-Küttner, C. (2018). Disappearance of Biased Visual Attention in Infants: Remediated Tonic Neck Reflex or Maturating Visual Asymmetry? *Perceptual & Motor Skills, 125*(5), 839-865.

Leo, I., & Simion, F. (2009a). Face processing at birth: A Thatcher illusion study. *Developmental Science, 12*(3), 492-498.

Leo, I., & Simion, F. (2009b). Newborns' Mooney-Face Perception. *Infancy, 14*(6), 641-653.

Liang, X., Liu, S., Wei, Y., Liu, L., Lin, L., & Yan, S. (2015). *Towards Computational Baby Learning: A Weakly-Supervised Approach for Object Detection*. Conference, https://doi.org/10.1109/iccv.2015.120

Liu, S., Xiao, N. G., Quinn, P. C., Zhu, D., Ge, L., Pascalis, O., & Lee, K. (2015). Asian infants show preference for own-race but not other-race female faces: The role of infant caregiving arrangements. *Frontiers in Psychology, 6*, 593.

Lockman, J. J., Ashmead, D. H., & Bushnell, E. W. (1984). The development of anticipatory hand orientation during infancy. *Journal of Experimental Child Psychology, 37*(1), 176-186. https://doi.org/https://doi.org/10.1016/0022-0965(84)90065-1

Lorenz, K. Z. (1937). The companion in the bird's world. *The Auk, 54*(3), 245-273.

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive psychology, 58*(4), 441-486. (Original work published Cognitive psychology)

Macfarlane, A. (1975). Olfaction in the development of social preferences in the human neonate. *Parent-infant interaction, 33*, 103-119.

Mareschal, D., Plunkett, K., & Harris, P. (1999). A computational and neuropsychological account of object-oriented behaviours in infancy. *Developmental Science, 2*(3), 306-317.

McFarland, D. (2014). *A Dictionary of Animal Behaviour* (2 ed.): Oxford University Press. https://doi.org/10.1093/acref/9780191761577.013.0209

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: toward an adaptive process account of successes and failures in object permanence tasks. *Psychological review, 104*(4), 686.

Needham, A. (1997). Factors affecting infants' use of featural information in object segregation. *Current Directions in Psychological Science, 6*(2), 26-33.

Needham, A. (1998). Infants' use of featural information in the segregation of stationary objects. *Infant Behavior and Development, 21*(1), 47-76.

Needham, A. (1999). The role of shape in 4-month-old infants' object segregation. *Infant Behavior and Development, 22*(2), 161-178.

Needham, A., Baillargeon, R., & Kaufman, L. (1997). Object segregation in infancy. *Advances in infancy research*.

Ng, W., Baillargeon, R., & Wilcox, T. (2007). 8.5-month-old infants use pattern but not color information to individuate objects in occlusion events

Park, J. C., Kim, D. S., & Nagai, Y. (2014). Developmental dynamics of RNNPB: New insight about infant action development (Vol. 8575 LNAI, pp. 144-153): Springer Verlag.

Pascalis, O., de Schonen, S., Morton, J., Deruelle, C., & Fabre-Grenet, M. (1995). Mother's face recognition by neonates: A replication and an extension. *Infant Behavior and Development, 18*(1), 79-85. https://doi.org/10.1016/0163-6383(95)90009-8

Perotto, F. S., & Alvares, L. O. (2006). Learning regularities with a constructivist agent*ACM.* Symposium conducted at the meeting of the Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems

Petrosino, G., & Parisi, D. (2015). A single computational model for many learning phenomena. *Cognitive Systems Research, 36-37*, 15-29. https://doi.org/10.1016/j.cogsys.2015.06.001

Piaget, J. (1954). The construction of reality in the child: New York: Basic Books.

Piaget, J., & Cook, M. (1952). *The origins of intelligence in children* [Electronic document]: New York : International Universities Press, [1952].

Pierris, G. (2017). Learning Robot Control Using a Hierarchical SOM-Based Encoding. *IEEE Transactions on Cognitive and Developmental Systems, Cognitive and Developmental Systems, IEEE Transactions on, IEEE Trans. Cogn. Dev. Syst.*(1), 30. https://doi.org/10.1109/TCDS.2017.2657744

Pirri, F. (2011). The well-designed logical robot: Learning and experience from observations to the Situation Calculus. *Artificial Intelligence, 175*, 378-415. https://doi.org/10.1016/j.artint.2010.04.016

Pissonnier, D., Thiery, J. C., Fabre-Nys, C., Poindron, P., & Keverne, E. B. (1985). The importance of olfactory bulb noradrenalin for maternal recognition in sheep. *Physiology & Behavior, 35*(3), 361-363.

Poggio, T. (1981). Marr's computational approach to vision. *Trends in Neurosciences, 4*, 258-262. https://doi.org/https://doi.org/10.1016/0166-2236(81)90081-3

R. A. Hinde, a., W. H. Thorpe, a., & M. A. Vince, a. (1956). The Following Response of Young Coots and Moorhens. *Behaviour*(2/3), 214.

Rochat, P. (1993). Hand-mouth coordination in the newborn: Morphology, determinants, and early development of a basic act. *Advances in psychology, 97*, 265-288.

Rosenblatt, J. S., Turkewitz, G., & Schneirla, T. C. (1961). Early socialization in the domestic cat as based on feeding and other relationships between female and young. *Determinants of infant behavior, 1*.

Ross, J. (2015). *Nervous system* [Electronic document]: Edinburgh, [Scotland] : Mosby, 2015.

Fourth edition.

Rovee-Collier, C., & Cuevas, K. (2008). The development of infant memory. In *The development of memory in infancy and childhood* (pp. 23-54): Psychology Press.

Rucci, M., & Victor, J. D. (2015). The unsteady eye: an information-processing stage, not a bug. *Trends in Neurosciences, 38*(4), 195-206. https://doi.org/https://doi.org/10.1016/j.tins.2015.01.005

Russell, M. J. (1976). Human olfactory communication. *Nature, 260*(5551), 520.

Sadeh, A., Acebo, C., Seifer, R., Aytur, S., & Carskadon, M. A. (1995). Activity-based assessment of sleep-wake patterns during the 1st year of life. *Infant Behavior and Development, 18*(3), 329-337.

Sai, F. Z. (2005). The role of the mother's voice in developing mother's face preference: Evidence for intermodal perception at birth. *Infant and Child Development, 14*(1), 29-50.

Salapatek, P., & Kessen, W. (1966). Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology, 3*(2), 155-167.

Salzen, E. A. (1962, 1962). Imprinting and fear

Salzen, E. A. (1978). Social attachment and a sense of security—a review. *Social Science Information, 17*(4-5), 555-627.

Schwartz, A. N., Campos, J. J., & Baisel Jr, E. J. (1973). The visual cliff: Cardiac and behavioral responses on the deep and shallow sides at five and nine months of age. *Journal of Experimental Child Psychology, 15*(1), 86-99.

Shultz, T. R. (2012). A constructive neural-network approach to modeling psychological development. *Cognitive Development, 27*, 383-400. https://doi.org/10.1016/j.cogdev.2012.08.002

Siegelmann, H. T. (2013). Review: Turing on Super-Turing and adaptivity. *Progress in Biophysics and Molecular Biology, 113*, 117-126. https://doi.org/10.1016/j.pbiomolbio.2013.03.013

Slater, A., Morison, V., & Rose, D. (1983). Perception of shape by the new-born baby. *British Journal of Developmental Psychology, 1*(2), 135-142.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science, 14*(1), 29-56.

Spelke, E. S. (1994). Initial knowledge: six suggestions. *Cognition, 50*(1), 431-445. https://doi.org/https://doi.org/10.1016/0010-0277(94)90039-6

Spelke, E. S., Breinlinger, K., Jacobson, K., & Phillips, A. (1993). Gestalt relations and object perception: A developmental study. *Perception, 22*(12), 1483-1501.

Squire, L. R., & Zola, S. M. (1996). Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences, 93*(24), 13515-13522.

Squire, L. R., & Zola-Morgan, S. (1991). The medial temporal lobe memory system. *Science, 253*(5026), 1380-1386.

Szenczi, P., Bánszegi, O., Urrutia, A., Faragó, T., & Hudson, R. (2016). Mother-offspring recognition in the domestic cat: Kittens recognize their own mother's call. *Developmental Psychobiology, 58*(5), 568-577. https://doi.org/10.1002/dev.21402

Szenczi, P., Banszegi, O., Urrutia, A., Hudson, R., & Farago, T. (2016). Mother-Offspring Recognition in the Domestic Cat: Kittens Recognize Their Own Mother's Call. *DEVELOPMENTAL PSYCHOBIOLOGY, 58*(5), 568-577.

Thelen, E., Corbetta, D., Kamm, K., Spencer, J. P., Schneider, K., & Zernicke, R. F. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. *Child Development, 64*(4), 1058-1098.

Turati, C., Bulf, H., & Simion, F. (2008). Newborns' face recognition over changes in viewpoint. *Cognition, 106*(3), 1300-1321.

Turati, C., Di Giorgio, E., Bardi, L., & Simion, F. (2010). Holistic face processing in newborns, 3-month-old infants, and adults: Evidence from the composite face effect. *Child Development, 81*(6), 1894-1905.

Turati, C., Macchi Cassia, V., Simion, F., & Leo, I. (2006). Newborns' face recognition: Role of inner and outer facial features. *Child Development, 77*(2), 297-311.

Valenza, E., & Bulf, H. (2011). Early development of object unity: Evidence for perceptual completion in newborns. *Developmental Science, 14*(4), 799-808. https://doi.org/10.1111/j.1467-7687.2010.01026.x

Valenza, E., Leo, I., Gava, L., & Simion, F. (2006). Perceptual Completion in Newborn Human Infants. *Child Development*(6), 1810.

Vervloed, M. P. J., Hendriks, A. W., & van den Eijnde, E. (2011). The effects of mothers' past infant-holding preferences on their adult children's face processing lateralisation. *Brain and cognition, 75*(3), 248-254.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in cognitive sciences, 6*(9), 376-382.

Wang, S. H., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition, 93*(3), 167-198.

Wang, S. H., Zhang, Y., & Baillargeon, R. (2016). Brief article: Young infants view physically possible support events as unexpected: New evidence for rule learning. *Cognition, 157*, 100-105. https://doi.org/10.1016/j.cognition.2016.08.021

Weingartner, H., Sitaram, N., & Gillin, J. C. (1979). The role of the cholinergic nervous system in memory consolidation. *Bulletin of the Psychonomic Society, 13*(1), 9-11. https://doi.org/10.3758/BF03334997

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition, 72*(2), 125-166.

Wilcox, T., & Baillargeon, R. (1998a). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive psychology, 37*(2), 97-155.

Wilcox, T., & Baillargeon, R. (1998b). Object individuation in young infants: Further evidence with an event-monitoring paradigm. *Developmental Science, 1*(1), 127-142.

Willig, C., & Stainton Rogers, W. (2017). *The SAGE Handbook of Qualitative Research in Psychology* (Vol. Second edition) [Book]. Thousand Oaks, California: SAGE Publications Ltd.

Wood, J. (2013). Newborn chickens generate invariant object representations at the onset of visual object experience. *Proceedings of the National Academy of Sciences of the United States of America*(34), 14000.

Wood, J., & Wood, S. (2018). The Development of Invariant Object Recognition Requires Visual Experience With Temporally Smooth Objects. *Cognitive Science, 42*(4), 1391-1406.

Wood, J. N. (2014). Newly hatched chicks solve the visual binding problem. *Psychological science, 25*(7), 1475-1481.

Wood, J. N. (2016). A smoothness constraint on the development of object recognition. *Cognition, 153*, 140-145.

Wood, J. N., Prasad, A., Goldman, J. G., & Wood, S. M. W. (2016). Enhanced learning of natural visual sequences in newborn chicks. *Animal cognition, 19*(4), 835-845.

Wood, S., Moriarty, K. M., Gardner, B. T., & Gardner, R. A. (1980). Object permanence in child and chimpanzee. *Animal Learning & Behavior, 8*(1), 3-9.

Worrall, J. (1989). Structural realism: The best of both worlds? *dialectica, 43*(1-2), 99-124.

Wu, D., Luo, Y., & Baillargeon, R. (2006). What object should appear in the window? 4-month-old infants' reasoning about inert and self-moving objects.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive psychology, 30*(2), 111-153.

Yeap, W. K. (2011). How Albot(0) Finds Its Way Home: A Novel Approach to Cognitive Mapping Using Robots (Vol. 3, pp. 707-721).

Yeap, W. K., & Hossain, M. (2019). *What is a cognitive map? - Unravelling its mystery using robots*. Auckland University of Technology, New Zealand.

Zentall, T. R. (2004). Action imitation in birds. *Learning & Behavior, 32*(1), 15-23.

Zlatev, J. (2001). The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines, 11*(2), 155-195. https://doi.org/10.1023/a:1011218919464

9 Appendix

Table 6 Summary of Empirical Findings

Citation	Age of Infants	Psychologists' Suggestions of Infants Capability
Salapatek and Kessen (1966)	8 Days	Prefer scanning vertexes
Valenza and Bulf (2011)	14 to 96 hours	Perceived stroboscopic motions as connected unit under occlusion
Valenza et al. (2006)	72 hours	Perceived stroboscopic motions as connected unit
Slater et al. (1983)	2 - 3 Days	Prefer shape discrimination to contour density
Leslie B. Cohen and Younger (1984)	6 weeks (1.5 months)	Orientation discrimination
Johnson and Nanez (1995)	2 months	No occlusion
S. J. Hespos and Baillargeon (2001b)	2 months	Containment
Aguiar and Baillargeon (1999)	2.5 months	Simple occlusion without height reasoning
Rovee-Collier and Cuevas (2008)	2 - 6 months	Memory tied with previous encountered experience
Spelke et al. (1993)	3 months	One whole object for both homogeneous and heterogeneous pattern of gestalt relation
Baillargeon and DeVos (1991)	3 months	No preferential difference for occlusion and height reasoning
Leslie B. Cohen and Younger (1984)	14 weeks (3.5 months)	Angular discrimination
Baillargeon and DeVos (1991)	3.5 months	Occlusion with height reasoning

Citation	Age of Infants	Psychologists' Suggestions of Infants Capability
S. H. Wang et al. (2004)	4 months	Reason hidden objects in occlusion and containment
S. J. Hespos and Baillargeon (2001a)	4.5 months	Height reasoning in occlusion event but not containment
Johnson and Nanez (1995)	4 months	Occlusion
Kellman and Spelke (1983)	4 months	Occlusion with continuous surface reasoning
Johnson et al. (2003)	4 months	Narrower occlusion gap
Bremner et al. (2005)	4 months	Narrower temporal gap facilitates occlusion responses
Bremner et al. (2007)	4 months	Orthogonal angle of trajectory tracking
Johnson and Aslin (1996)	4 months	Object unity facilitated by background with texture, edge alignment and relatability
Amy Needham (1999)	4 months	Utilize shape differences more than pattern and colour to constitute boundaries
Johnson et al. (2000)	4 months	Gestalt principle of figural goodness
Wu et al. (2006)	4 months	Objects occupy a space
Amy Needham (1997)	4.5 months	No details of features, shapes of the objects could be used to segregate object
Amy Needham (1998)	4.5 months	Reason object segregation with simpler shapes
Wilcox and Baillargeon (1998b)	4.5 month	Use featural information to individuate object but with lower tolerance with occlusion period
Teresa Wilcox (1999)	4.5 month	Use size and shape only to distinguish object
Csibra (2001)	5 months	Affect by habituation sequence to gestalt principles in occlusion
Baillargeon, Spelke, and Wasserman (1985)	5 month	Object Permanence
Johnson et al. (2003)	6 months	Wider occlusion gap

Citation	Age of Infants	Psychologists' Suggestions of Infants Capability
Spelke et al. (1993)	5, 9 months	Homogenous pattern as one object, heterogenous pattern as different objects of gestalt relation
Gronqvist et al. (2006)	5, 7 months	Horizontal trajectory tracking
Bertenthal, Gredeback, and Boyer (2013)	5 - 7 months	Prefer kinetic movement but could be trained to track other types of movement
Bertenthal, Longo, and Kenny (2007)	5, 7, and 9 months	Track best with kinetic movements and become more capable to track other form of movements when grow older
Amy Needham (1997)	6.5 - 7.5 months	Use featural information to individuate object
S. J. Hespos and Baillargeon (2001a)	7.5 months	Height reasoning in containment event
T. Wilcox and Baillargeon (1998a)	7.5 month	Use featural information to individuate object
Teresa Wilcox (1999)	7.5 month	Use pattern to distinguish object
Amy Needham (1998)	7.5 months	Reason object segregation with curved / more complex shapes
Csibra (2001)	8 months	Does not affect by habituation sequence of illusionary occluder
Amy Needham (1997)	8 months	Prefer spatial pattern to analyse object boundary, could use physical rule to reason object unity, failed to use pattern to segregate objects
Ng, Baillargeon, and Wilcox (2007)	8.5 month	Use pattern not colour to distinguish object
Gronqvist et al. (2006)	9 months	Both horizontal and vertical trajectory tracking

Citation	Age of Infants	Psychologists' Suggestions of Infants Capability
Xu and Carey (1996)	10 months	Expect an object appeared behind an occluder but fail to distinguish the same toy with different colour and texture
Teresa Wilcox (1999)	11.5 month	Use colour to distinguish object
Xu and Carey (1996)	12 months	Object individuation