

**The Color Identification Puzzle: A
Proposed Solution and its
Implications**

by

Joel Salon

**A dissertation submitted to the Graduate Faculty
in Psychology in partial fulfillment of the
requirements for the degree of Doctor of
Philosophy, the City University of New York**

2010

Copyright 2010
Joel Salon
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Psychology.

Date: _____

Katherine Nelson
Chairman of Examining Council

Date: _____

Maureen O'Connor
Executive Officer

Katherine Nelson

Laraine McDonough

Bruce Homer
Supervising Committee

Abstract

The Color Identification Puzzle: A
Proposed Solution and its
Implications

by

Joel Salon

Adviser: Professor Katherine Nelson

That children have difficulty learning to correctly identify the color of objects is well established. The question is why. This research is based on the hypothesis that the brain monitors the environment to identify cues that will reliably enable it to predict what will happen next and that the primary learning process used by the brain is to compare its prediction of what is going to happen next in the world to what, in fact, actually happens next. As a result, aspects of a child's environment that do not reliably enable a child to predict what is going to happen next or otherwise do not assist the child to understand his environment or participate in activities with others are not attended to by the brain. It is argued that color is just such an "ignorable" aspect of the environment of most children. To test these hypotheses, 27 pre-school children who were unable to consistently correctly identify the color of blue, red and yellow objects were divided into two groups. The experimental group was taught using a training methodology designed to engage the hypothesized prediction process. The control group was taught using a training methodology that did not engage the prediction process. The results demonstrated that the experimental training methodology was more effective in teaching children to correctly identify the color of objects than was the control training methodology.

Acknowledgments.

This dissertation would not have been possible without the patience and support of my advisor, Katherine Nelson, my wife, Carolyn, my daughter, Heather and my assistant, Lolita Ali.

Heather was instrumental in designing the materials, finding participants for and carrying out a successful pilot study. Without Carolyn's tireless efforts in finding participants and testing and teaching the participants, this dissertation would not have come to fruition. Lolita's patience with seemingly endless revisions was and is appreciated.

Most importantly, Katherine Nelson's open minded support for my second year connectionist computer modeling project paved the way for the ideas motivating this dissertation and her functionalist approach to word learning is the other foundation upon which this dissertation rests.

Thank you also to the other members of my committee, Laraine McDonough and Bruce Homer, for sharing their advice and to Herbert Saltzstein and Patricia Brooks, who took the time to be my readers.

Table of Contents

	<u>Page</u>
Chapter 1. Introduction and Background for the Theory of Prediction, Filtering and Inhibition Processes in Learning and Memory.	1
Introduction.....	1
General Considerations.....	3
Evolutionary Considerations.....	5
Classical Conditioning and Operant Conditioning.....	10
Movement; Perception and Reafference Theory.....	12
Statistical Learning.....	13
Connectionist Theory.....	15
Event Knowledge Theory.....	17
Neuroscience.....	21
Neoconstructivism.....	22
Conclusion.....	22
Chapter 2. Learning Color Words.....	24
The Early Word Learning Literature.....	24
The Color Word Literature.....	25
Chapter 3. Experiment I – Learning Colors.....	36
Research Questions.....	36
General Description.....	37
Participants.....	38
Materials.....	39
Procedure.....	39
Pre-Test.....	39
Experimental Training Procedure.....	41
Control Training Procedure.....	42
Post-Test.....	42
Results.....	42

Chapter 4 Experiment II – Learning Shapes	49
Method	49
Research Questions.....	49
Participants.....	49
Materials	49
Pre-Test	49
Experimental Training Procedure	50
Control Training Procedure	50
Post-Test	50
Results.....	50
Chapter 5. Discussion.....	52
Experiment I.....	53
Experiment II	56
Conclusions and Implications	58
Endnotes	63
References.....	65

List of Tables

	<u>Page</u>
Table 1 – Pre-Test Results	44
Table 2 – Pre-Test Results Combined for Both Groups	45
Table 3 – Post-Test Results.....	46
Table 4 – Post-Test Minus Pre-Test Results.....	47
Table 5 – Post-Test Versus Pre-Test Results Comparison of Mean Increase in Total Scores for Participants Under Four Years of Age.....	48
Table 6 – Comparison of Mean Increase in Total Scores for Participants Scoring Below the Mean Total Score on the Pre-Test	49
Table 7 – Comparison of Increase in Mean Scores for Colors Versus Shapes	52

Chapter 1. Introduction and Background for the Theory of Prediction, Filtering and Inhibition Processes in Learning and Memory.

Introduction

It has long been known that while most children starting at around age two begin to learn words rapidly and easily, children have difficulty learning to correctly apply common color words which they can enunciate such as red, blue and yellow to correctly name the color of objects (color naming) and that this difficulty persists for many children as old as four years of age and even some children as old as five years of age still have difficulty in color naming. Darwin was among the first to comment on this, initially speculating that his own children were color blind because of the difficulty he had in teaching them to correctly identify colors. Subsequently, numerous investigations of this phenomenon have been carried out and a variety of explanations have been proposed.

In this dissertation I propose that the explanation for the difficulty children have in learning to correctly apply color words is that young children have learned to ignore and, therefore, do not attend to the color of an object because the color of an object does not reliably predict anything meaningful about the object and its relations or its functions. I report a study testing the hypothesis that the color naming difficulty that many children have is related to basic properties of brain processing.

Specifically, the hypothesis is that learning and memory are dependent on the following neural processes:

The brain (1) constantly monitors its environment in order to identify environmental cues (a “Cue Suite”) that will enable it to predict what will happen next,

(2) predicts what will happen next and (3) compares its prediction of what will happen next after the occurrence of a Cue Suite to what, in fact, does happen next (the “Prediction Process”). The match or mismatch, as the case may be, between the brain’s prediction of what will happen next and what actually happens next drives the learning process. Since the brain is bombarded with stimuli from the environment but does not have unlimited processing capacity, an important corollary of the Prediction Process is that the brain learns to ignore stimuli that it has come to expect are not reliable predictors of what will happen next (the “Filtering Process”).¹ It is likely as well that the brain monitors the environment for Cue Suites that are important for survival such as facial and body posture and movements that predict the intentions of others.

A training procedure designed to engage the Prediction Process was used to teach children in the experimental group to correctly name the color of objects and the results of the experimental training procedure were compared to the results achieved by a control group receiving closely similar training that did not engage the Prediction Process. The success of the experimental training procedure (reported in Chapter 4) supports the Prediction Process and Filtering Process hypotheses with respect to learning in everyday situations. Its success may have additional implications for teaching abstract concepts in general.

Both the motivation and support for these hypotheses is presented in more detail in this chapter and, in particular, it is shown that these processes are consistent with a wide range of neuro-cognitive, cognitive and developmental theories and experimental work. While certain of the processes are explicitly relied upon in these theories, e.g.,

reafference theory, they are implicit in others. The suggestion is that by explicitly recognizing these processes as fundamental to all learning, many of the extant theories and experimental work can be harmonized. In particular, how the Prediction Process and Filtering Process provide an explanation for the difficulty many children have in learning to correctly name the color of objects is presented in this chapter.

General Considerations

At first blush, human cognition and the behavior it generates appear much too complex and unpredictable to be governed by a few simple processes as is suggested above. However, it is also remarkable that the recursive application of very simple rules often times generates behavior so complex and seemingly unpredictable as to be called chaotic (Gleick, 1987). Such behavior is described as chaotic for the reason that very small differences in the assumed or measured initial conditions of the system results over time in very different outcomes.

Life could not come into being or remain in existence in a truly chaotic universe. The theory of evolution requires an environment where some regularity or predictability exists – at least on a statistical basis – so that variations conferring survival advantages for an organism will more often than not be selected over time. Thus, many chaotic systems – take the weather as an example – exhibit sufficient regularities so that the behavior of the system over short periods of time and the general behavior of the system over long periods of time, while not precisely predictable, are predictable within a well-defined range. Everyday observation suggests that human cognition and behavior have some of the characteristics of chaotic systems in that what a person will do at any

particular point in time is not precisely predictable and becomes less so the further out in time one tries to predict what a person will do. However, the range of reaction of a person given a described set of circumstances is generally predictable. Similarly, while most healthy infants are able to successfully reach and grasp objects at around four months of age, a study of several infants showed that while they all achieved this predictable developmental milestone at around the same time, their developmental paths varied depending upon, among other things, their temperament and physical attributes.² (Thelen et.al., 1993).

The complexity of the mind is nothing if not daunting. It has been estimated that the human brain contains some one thousand billion neurons, resulting in a network with over one thousand trillion connections in the human cerebral cortex (Lowenstein, 1999). Indeed, it is certainly possible, as some have claimed, that the human mind is not powerful enough to understand itself (Penrose, 1989). But the history of science belies such pessimism as seemingly unrealistic simplifications of complex phenomena have time and time again led to the discovery of powerful underlying principles. Thus, Newton in formulating his laws of motion considered an idealized state without friction. Yet, Newton's laws of motion and similar principles have then in turn been employed to enhance our understanding of the phenomena excluded from consideration in the idealized state such as friction.

The underlying assumption of this dissertation is that the Prediction Process, Filtering Process and Inhibition Process – see endnote 2 to this chapter – can explain and harmonize most of the properties of learning both in humans and animals that have been

firmly established by Pavlov, Skinner and Reafference Theory, among others. Further, the power of these processes undercut the need to resort to nativist “explanations” of human learning such as the so-called language organ proposed by Chomsky (1965). An understanding of the power of these processes coupled with the fact that the brain will quickly generalize from small samples can also help explain (i) what often appears to be superstitions and irrational seeming beliefs and behavior, as well as (ii) the biases people exhibit such as the confirmation bias where people attend to facts that confirm their existing opinion and ignore facts (the Filtering Process) that are inconsistent with their existing opinion, and (iii) that generally people make choices and behave in ways that are not strictly logical (Tversky & Kahneman, 1973 and 1974). Finally, although not explored here (but see neuro-cognition considerations infra), there is substantial evidence that perception is itself dependent upon Prediction Process learning (Dambacher, et.al., 2009; Raichle, 2010).

Evolutionary Considerations

As Darwin was one of the first, if not the first, to comment on the difficulty children have in learning to correctly identify the color of objects, it is perhaps not out of place to begin with an evolutionary “just so” story to begin to establish the plausibility of the Prediction Process being the main driver of learning.

Multi-cellular organisms (with some oddball exceptions) are classified either as plants or animals. Although biologists do not distinguish between the animal and plant kingdoms based upon whether an organism is self-mobile or not, it does happen to be the case that most plants (perhaps all) are not self-mobile while most organisms classified as

animals are self-mobile. Plants are able to survive without moving from place to place, because they are able to convert the energy in sunlight into a form of energy that can be stored and then used as needed, together with the resources of the air, ground and water available to them where they grow to sustain life and reproduce. Animals cannot directly utilize the energy in sunlight and the ability of an animal to move through its environment in a coordinated manner in order to collect energy in the form of food as well as avoid danger and mate is essential for its survival.

It is also the case that most, if not all, multi-cellular organisms that are self-mobile have neuron assemblies that coordinate and direct movements and specialized neuron assemblies for gathering information about the organisms' environment. On the other hand, plants do not move and do not have neuron assemblies. Evolutionary change is driven by variations that confer on individuals a statistical advantage over their conspecifics in producing offspring. That is to say, the organism with a variation that enables it to correctly guess what will happen next more often than its conspecifics will be more likely to survive and thus produce offspring all other things being equal. Thus, it is plausible from an evolutionary perspective to suggest that neuron and neuron assemblies were selected in the first instance to gather information about the environment generally and to facilitate goal directed movement because the ability to monitor the environment for possible threats and opportunities and to move to avoid threats and take advantage of opportunities conferred a survival advantage.

It has become widely acknowledged that perception itself often relies upon prediction. In the Dambacher, et.al. (2009) article entitled “Event-Related Potentials Reveal Rapid Verification of Predicted Visual Input”, the authors state:

We regard such anticipatory activity as top-down predictions engaging lower-level areas involved in sensory processing to grant fast and smooth perception of forthcoming stimuli. Given that the quantity of feedback connections to primary sensory areas even outnumbers pure feedforward input the interplay of top-down and bottom-up flow appears as a major principle of perception. (p.2)

Moreover, as noted in Raichele (2010):

Researchers have known for some time that only a trickle of information from the virtually infinite flood in the surrounding environment reaches the brain’s processing centers. Although six million bits are transmitted through the optic nerve, for instance, only 10,000 bits make it to the brain’s visual-processing areas, and only a few hundred bits are involved in formulating a conscious perception – too little to generate a meaningful perception on their own. The finding suggested that the brain probably makes constant predictions about the outside environment in anticipation of paltry sensory inputs reaching it from the outside world. (p.47)

Similarly, movement is an activity that must be computed on-line in response to changes in the topology of an organism’s environment as well as changes in and the threats and opportunities presented by the organism’s environment. The organism must organize the muscle movements required to attain the desired goals and over time through feedback learn to effortlessly control its bodily movements. Indeed, the difficulty of navigating over an irregular topology and keeping one’s balance on two limbs instead of four limbs while freeing the other two limbs to simultaneously engage in separate tasks and motions ratchets up the need for accurate prediction. That the motor

areas of our brains are generally one step ahead of us, that is, it is predicting what will come next (Praamstra, et.al., 1995) becomes evident to us when we step off a stair while distracted incorrectly anticipating that it is the next to last stair when in fact it is the last stair. The resulting jolt is an example of the unpleasant consequences that result from a failure to accurately predict what will come next (Freeman, 1990).

The proper organization of muscle movement is not a task that can be easily performed through logical computations or information processing because, among other things, of the many degrees of freedom that are involved (Spoms & Edelman, 2008) leading to a combinatorial explosion if logical analysis is the sole method available to “compute” the right “answer”. The painfully slow development of a robot that can move through the environment and interact with and manipulate the objects in its environment makes it clear that movement is a difficult digital computing problem, but one that many organisms with little or no brain can solve. On the other hand, computers are far more adept than humans at using logic to analyze and manipulate data and even create mathematical proofs (Simon & Newell, 1958). Taken together these observations coupled with the principle that evolution is conservative, meaning that it is easier to use or modify structures and processes that already exist, suggests that the process of prediction that enables coordinated movement (activity) and perception were co-opted by evolution for use as a general learning process (Diamond, 2000; Dambacher, et.al., 2009).

When one contrasts the controlled environment that is necessary for computers to function at all with the moment-to-moment challenges that biological organisms must deal with successfully in a semi-predictable but ever changing environment in order to

survive, then one is forced to look at the “logic” of evolution for an understanding of human cognition and not to the logic of digital based information processing. The “logic” of evolution is survival for reproduction and the argument here is that the most useful ability an organism can have as it relates to survival is the ability to be able to predict what is likely to happen next in its environmental niche with respect to (i) the inanimate objects and animate organisms that populate its environmental niche, (ii) its actions on such inanimate objects and animate organisms and (iii) the actions of other organisms and their likely consequences.

Thus, the basic processes underlying animal and human cognition, namely, attention, perception, association, memory, inhibition and sensitization are all used by complex organisms either singly or in combination with one another to enable the organism to predict what is going to happen next, which, if one is intent on surviving, is the most important information one can have. If one accepts that neuron networks evolved to support perception and control of movement and that the coordination of the organism’s movement through the environment and perception itself both rely, at least in part, on the Prediction Process, then one comes naturally to the hypothesis that the Prediction Process must also play a key role in cognition generally.

Moreover, that the Prediction Process together with the Filtering Process and Inhibition Process and, of course, memory are the core cognitive processes used by organisms in learning is supported by three of the most robust psychological phenomena of animal and human behavior – classical conditioning, operant conditioning and

reafference theory – and is consistent with most developmental theories concerning the development of early cognitive abilities.

Classical Conditioning and Operant Conditioning

Classical or Pavlovian conditioning has been described as follows:

Pavlovian conditioning is presumed to involve learning about relationships between stimulus events in the environment. Pavlov (1927) argued that temporal contiguity between an initially neutral conditional stimulus (CS) and a motivationally significant, unconditioned stimulus (US) is sufficient to produce conditioning. However, we now know that mere temporal CS-US pairing does not always result in conditioning. Whether conditioning occurs also seems to depend on whether there are occurrences of CS or US in the absence of the other, i.e., on the “contingency” between CS and US. (Gibbon, Berry, Thompson, 1974; Resorla, 1968).”

Dess et.al., 1986, pp. 31-32

What is being suggested here is that mere association does not lead to learning. If, for example, Pavlov always wore a white lab coat when in the presence of his dogs, but did not always ring the bell when in the presence of his dogs and if food always followed the ringing of the bell, then although Pavlov and his white lab coat were always and consistently associated with food, his appearance in a white lab coat did not unambiguously predict food while the ringing of the bell did. Since only the ringing of the bell made a difference, that is, for the dog, the ringing of the bell and not the white lab coat predicted food, only the ringing of the bell would constitute an environmental cue that would be remembered by the dog in relation to the appearance of food. Therefore, the ringing of the bell and not the appearance of Pavlov in his white lab coat would elicit the unconditioned response. (Joseph Glick, personal communication).

While classical conditioning involves an unconditioned stimulus inducing an involuntary response, that is, a stimulus-response relationship, operant conditioning is where a voluntary response is followed by a reinforcing stimulus, that is, a response-stimulus relationship. Nonetheless, operant conditioning takes place because the response or behavior of pressing a bar, for example, results in a food pellet being released at least some of the time. Again, it is the contingent relationship – or the predictable relationship – that drives the learning process (Skinner, 1969). The organism is predicting that certain actions will result in certain outcomes and those actions that do in fact result in desired outcomes are the actions the organism learns and performs. That is not to say the organism is always correct and humans, in particular, are notorious for overgeneralizing from small samples. Thus, are born superstitious.

Indeed, the fact that a schedule of intermittent reinforcement is not only sufficient for learning, but also results in learning that is the most difficult to extinguish is further evidence that complex organisms including humans are particularly sensitive to environmental cues that predict what happens next even where the relationship is only a statistical relationship and not an absolute unvarying relationship. Given the complex nature of the environment and that Darwinian evolution is based upon a process of statistical selection, this makes perfect sense, since a set of organisms carrying a particular set of genes enabling them to be better predictors need only be more often right on average than a different set of organisms that do not carry the particular set of genes in order for the first set of organisms to be the likely winners of the evolutionary lottery.

Movement; Perception and Reafference Theory

Reafference Theory, which dates back to Von Helmholtz's work in the 1860s on the physiology of the eye (Wolf, 1999) provides further support for the Prediction Process. Reafference Theory has provided evidence that movement requires that the brain not only constructs action sequences in advance of movement, but that the brain also predicts the sensory input that will result from the planned action sequence and then verifies through sensory reafference including proprioception that the planned action sequence has resulted in the sequence of sensory inputs that were predicted by the brain (Freeman, 1990).

Freeman describes the process as follows:

...perception instead begins with an internally generated neural process that prepares the organism to seek out future stimuli in the outside world. What actually occurs is the emergence of a motor action pattern—a command, if you like—in the nervous system, which is then fed out to the muscles and glands through the brainstem. It is important to note, however, that most of this outflow remains within the brain itself where, redirected to the receiving areas of the cortex, it alerts the sensory systems to expect the sensory consequences of motor action (Hemholtz, 1879). Through this self-reflexive update, now known as corollary discharge (von Holst and Mittelstaedt, 1950), by the process of reafference (Sperry, 1950), the brain actively grasps for the consequences of action into the environment.

Thus, Paillard, arguing in this tradition, states that one should look at the brain as an “anticipating machine” (Paillard, J., 1999). Similarly, Dambacher et.al. (2009) argue that “human information processing depends critically on continuous predictions of upcoming events, . . .” (p.1). While Grossberg (1998) argues that consciousness itself depends on processes that:

include the learning of top-down expectations, the matching of these expectations against bottom-up data, the focusing of attention upon the expected clusters of information and the development of resonant states between bottom-up and top-down processes as they reach an attentive consensus between what is expected and what is there in the outside world. (p.1)

Statistical Learning

If the Prediction Process drives learning and is the basis of cognition, then since precise prediction is impossible in practice complex organisms must rely on probability estimates. Probability estimates are in turn based upon frequency. Thus, complex organisms must be able to keep track of event frequencies if the Prediction Process hypothesis is correct, and there is evidence that they are able to do so.

For example, Maye, Werker & Gerken, (2002) showed that six-months old infants were sensitive to statistical information in speech sounds. One group was exposed for 2.3 minutes to a uni-modal distribution and the other to a bi-modal distribution of the sounds “ta” and “da” for the same time period. In the bi-modal distribution the stimuli near the end-points of the “ta”–“da” phonetic continuum occurred more frequently. In the uni-modal distribution the stimuli in the middle of the “ta” – “da” phonetic continuum occurred more frequently. After this short exposure, the infants presented with a uni-modal distribution were unable to distinguish between the sounds while those exposed to a bi-modal distribution were able to do so. The authors suggested that their study demonstrated a domain-general probability learning mechanism that could account for both the ability of infants to learn the contrasts relevant to their native language and account for the fact that many non-native contrasts, that is, contrasts that are not relevant

for discriminating speech in one's native language that are discriminable early in infancy are no longer discernable in later infancy (Best et.al., 1995; Werker & Tees, 1984). This loss of the ability to distinguish phonetic contrasts that do not make a difference is a clear example of the Filtering Process at work.

Subsequently, Saffran, Aslin and Newport (1996) demonstrated that 8-month-olds can sequence an artificial language composed of trisyllabic nonsense words that are solely defined by their probabilities of co-occurrence after only a two-minute familiarization period. The "words" in this study were "defined" by a sequence of sounds and not by pauses, prosodic cues or other cues. Their study clearly demonstrates that statistical learning emerges early and is quite powerful. [See, also, Saffran, (2010) for a summary of recent confirming research and, also, Elman's connectionist network that discovers word boundaries by predicting the next sound (Elman, 1990)].

Moreover, connectionist computational models, using statistical learning algorithms have presented a convincing simulation of the acquisition of the past tense in English including the phenomenon whereby young learners first correctly use verbs whose past tense is irregular such as go/went that were initially learned by them, but then over-generalize the use of the "ed" ending and produce endings such as go/goed. Further, both the toddlers and the connectionist simulation eventually learn to correctly use irregulars (Plunket & Marchman, 1991).

Since the hypothesized Prediction Process is probabilistic, it is likely that early experience may lead the organism to mistakenly-generalize the applicability of a heuristic that is applicable in only certain contexts. So, for example, a toddler will notice that if he

asks for more juice, the level of the juice in the glass will invariably be higher after his mother pours more juice into the glass. Thus, the child learns that the higher the level of the liquid in the glass, the more liquid there is in the glass. When given the traditional Piagetian liquid conservation task which requires the child to attend to both the height and width of the container, the challenge is for the child to learn to inhibit this well entrenched heuristic.³

Connectionist Theory

As noted earlier, connectionist theory also supports the Prediction Process. Connectionist theorists attempt to discover how multiple connected neuron assemblies interact to produce cognition by designing simplified models of neural networks and then running simulations on them using simple rules to govern their evolution over time. (See, for example, Parallel Distributed Processing (Rumelhart, McClelland and the PDP Research Group, 1986), the seminal work in this area.)

Connectionist networks are made of nodes, simple processors meant to simulate neurons, and connections between the nodes which are meant to simulate synaptic connections between neurons. One form of connectionist network that has been successful in simulating learning without relying upon symbolic representation and rule based analysis is a network that has a set of input nodes – simulating sensory neurons – a set of middle nodes often described as hidden nodes – simulating subconscious processing – a set of recurrent nodes – simulating memory – and a set of output nodes – simulating the response to the input or consciousness – and a bias node – simulating the fact that typically neurons have threshold characteristics.

The strength of the connections in the system are set randomly to begin with. The goal of the training regime is to adjust the strength or weight of the connections so that the output nodes produce the desired pattern of activation – simulating the desired state or behavior – for a particular pattern of activation at the input nodes – simulating the environmental sensory input. The network is “trained” (the network learns) by adjusting the strength of the connections between the input nodes and the hidden nodes, between the hidden nodes and the output nodes, between the bias node and the hidden nodes and output nodes. Generally, the strength of the connections between the recurrent nodes and the hidden nodes are not adjusted as the recurrent nodes feed back to the hidden nodes the prior state of the system. One of the more successful algorithms for adjusting the strength of all of these connections is called the back-propagation algorithm.

The back-propagation algorithm works by comparing the activation at the output nodes to the desired activation and uses a complex formula to adjust the connection strengths. For example, the back-propagation algorithm will weaken the connections between an output node that should have zero activation but on a particular training run is showing a positive activation with the hidden nodes that are, as it were, contributing to the positive activation, that is, contributing to the error between what the system’s output state and what, in fact, the output state should be.

The back-propagation algorithm has been criticized as not being biologically plausible in that the neural networks in the brain do not appear to back-propagate signals and because it assumes the brain “knows” the answer in advance. (But see, Hawkins, 2004). However, if the brain is predicting what will happen next before it happens, then

the neuron connections that predicted what is going to happen next may be adjusted based upon how closely the prediction is to what actually occurred by a process not now understood.

In particular, a connectionist model of the type described above was trained successfully with the back-propagation algorithm to predict a sequence of events (Salon unpublished paper, 2000). The sequence of events was designed to model the occasion of eating a meal for a toddler. The goal of this network was to predict the next event in the sequence, for example, that after the toddler is put in the high chair, the toddler's bib is put on and so forth. While the sequence of events was invariant, the type of food that was put on the high chair in front of the child following the event of drinking juice would be randomly varied. However, even though the foods presented could be different, the system had to predict the next event – the toddler eats – regardless of the type of food input into the system. (Each type of food was represented by a different and unrelated pattern of activity on the input nodes, but the connectionist system was required to convert these unrelated input activation patterns into the same output activation pattern representing the next event in the sequence, that is, eating.)

This particular connectionist simulation was designed to provide support for event knowledge theory and, in particular, slot-filler categorization. (See, for example, Nelson, 1986).

Event Knowledge Theory

In Event Knowledge (Nelson, 1986), Katherine Nelson explored the seeming disconnect between children's performance on experimental tasks and their performance

on everyday tasks (when the latter would seem to require the same or even more complex cognitive abilities as the former) by hypothesizing that representations of events constitute a basic form of a child's knowledge of everyday life situations and that the more abstract structures of classification, memory, logical problem solving and language abilities are, in part, derived from what Nelson called generalized event representations.

As Nelson noted, events have structures. At the simplest level, events unfold in a temporal sequence which may, in turn, be determined wholly or partially by causal mechanisms or logical constraints. Events can also be hierarchically organized. For example, the event of going shopping may have embedded within it the familiar event of (at least in winter) first putting on clothes for the outside, the familiar event of driving, the event of shopping in the store itself, the event of paying at a cash register and so on.

But first and foremost there is a temporal sequence to any event and that sequence is in large measure invariant because of real world constraints. Thus, the driving event requires that the vehicle (which may vary) must be opened (motorcycles excluded), then the participants in the event must seat themselves, then the vehicle must be started, etc. In order for a child to cooperatively participate in this event she must, at a minimum, be able to predict from the current action or state the next action to be taken in the event.

Although everyday events have an invariant or semi-invariant structure, certain portions of that structure are indeed invariant in all respects while other portions can display variability. However, generally that variability is itself constrained. Thus, the event of eating for a young child can be conceived as having the following structure. At a given time of day (breakfast, lunch, dinner), a caretaker (mother, father, other person)

puts the child in a highchair, the caretaker then puts a bib on the child and puts utensils and tableware on the highchair (perhaps different bibs, different tableware and perhaps in reverse order), the caretaker then provides food and drink (various items often regularly related to whether the event is breakfast, lunch or dinner), food is placed in child's mouth (food is required to be chewed, food is not chewed if liquid or sufficiently preprocessed), food is swallowed, tableware and utensils are taken away, the child's bib is taken off, and finally the child is taken out of the highchair.

As can be seen, the sequence of sub-events is relatively invariant, some of the actions are invariant, e.g., the child is put in the highchair, and some have variable elements, e.g., the various types of food that may be eaten. If one conceptualizes each sequence unit as a slot, then certain slots, while both reliably following certain other slots and preceding certain other slots, can be filled with a variety of different actors, objects or actions.

Nelson (1985) proposed that this property of substitutability in everyday event structures leads to the creation of slot-filler categories. Such context-defined categories are different from abstract "logical" categories. Thus, the logical hierarchical structure of the category for food transforms the context bound, slot filler categories developed by the child for food for breakfast, food for lunch, etc., into food as anything that can be eaten. She noted that many familiar object categories are found in everyday events experienced by children such as foods that are eaten for breakfast. Nelson claimed that the achievement of a logical (non-context dependent) superordinate category such as food is based upon and develops over time out of the child's early grouping of slot-filler food

items together. Nelson pointed out that similarity cannot explain the formation of the logical category of food since foods often bear little perceptual similarity to one another (e.g., milk and carrots). Moreover, although reliable contiguity is necessary for the event to have a coherent associative structure from which information can be extracted; contiguity alone cannot explain the formation of the logical category of food inasmuch as milk and carrots are not necessarily present in the same context at the same time.

The salience of slot filler categories for children as opposed to taxonomic or logical categories has been demonstrated using a variety of research paradigms including a free recall task where subjects are presented with superordinate labels and asked to produce category members, a word association task where subjects are asked to respond to a stimulus word with the first word that comes to mind and a forced-picture-choice task where subject must choose which one of two pictures goes best with a target picture. (See Lucariello, Kyratzis and Nelson, 1992). While the salience of slot filler categories has been challenged by some researchers with Krackow and Gordon (1998) arguing that the results supporting the slot filler hypothesis can be explained by mechanisms of association and typicality, the slot filler hypothesis is consistent with the Prediction Process, can be simulated in a connectionist network (see below) and can also explain how children learn to participate in everyday event sequences.

Tomasello and Brooks (1999) make a similar argument with respect to the learning of grammar. Their claim is that early syntactic development emerges from “. . . children’s understanding of the various scenes that make up their lives, including both dynamic events and relatively static states of affairs” (p.161).

The author's neural simulation did in fact "learn" to predict the next event after the presentation of food. Moreover, the connectionist network succeeded at this task by transforming the different activation patterns at the input nodes for each of the foods into substantially the same activation pattern at the hidden layer, that is, it developed, if you will, a slot filler category.

Thus, while the goal of the network was to predict the next event in the sequence and not to categorize, categorization was achieved, that is, categorization constituted an emergent self-organizing property of the Prediction Process.

Neuroscience

In On Intelligence, Jeff Hawkins (2004), a computer scientist and neuroscientist, proposed that the brain is not at all similar to a digital computer but rather that it is a memory prediction system. Hawkins came to this conclusion solely from studies regarding the organization of the cerebral cortex. The main insight was Mountcastle's (1978). He argued that given the uniformity in structure of the neocortex that each of its sections—visual, motor, hearing, etc. including intelligence—must work on a common principle or algorithm. In 1991, Felleman and Essen described the hierarchical organization of the visual cortex and Hawkins applied it to explain intelligence generally and the cortical processes that support the prediction and learning of sequences of patterns.

The details as he notes are likely to be incorrect. However, what is significant is that he comes to the same conclusion—the engine of cognition is prediction—from an

entirely different perspective than the one presented here. See also Grossberg (1999), Paillard (1999), Diamond, (2000) and Dambacher et.al. (2009).

Neoconstructivism

In reaction to nativist claims that seemed to leave little room for either learning or development, Neoconstructivism, a new paradigm spurred by the success of connectionist modeling, has developed. Its core principles as enunciated by Newcombe (2010) are all consonant with the principles and analysis discussed above, to wit: (i) the mechanisms for learning and development must be plausible from an evolutionary perspective; (ii) experience expectancy is a key concept; (iii) the world is richly structured with correlations that support experience-expectant learning; (iv) humans learning is supported by a domain general statistical learning ability; (v) the ability to select among and/or integrate multiple environmental cues; and (vi) active movement plays a key role in learning and development.

Conclusion

Psychologists have long used the language of expectation and prediction in explaining behavior. The experience and evidence supporting classical and operant conditioning establish beyond doubt that at least certain behaviors are driven by expectation or prediction. Jeff Hawkins, who came to his version of the Prediction Process from a strictly neural science point of view, also acknowledged that this is not a new concept. See also Paillard who said that one should look at the brain as an “anticipating machine” (Paillard, 1999). The connectionist models, particularly the recurrent prediction network described above, are “existence proofs” showing, at the very

least, that the brain's processes need not be based on the logical manipulation of symbols or representations and that knowledge does not have to be built-in.

The suggestion here is that the Prediction Process together with the Filtering Process and Inhibition Process (all of which rely upon memory) should be considered as the primary drivers of intelligent adaptive behaviors and that such processes can explain, together with effects of emotion on these processes and certain very limited genetically hardwired behaviors, how most learning occurs.

The observation that culture and context play such a large role in cognition is also consistent with the model of cognition described here because it is the context that provides the stimuli that enables the organism to predict (not logically compute) what is going to happen next. Likewise, Gibson's concept of invariance and affordance are consistent with the model of cognition described here as is the neocortex model described in detail by Jeff Hawkins and the principles of Neoconstructivism. Similarly, Nelson's work on event knowledge and function and meaning is consistent with the proposed model.

Had the experiment described here failed it would not, of course, disconfirm the proposed cognitive model nor does its success confirm it. There is probably no one experiment that could be devised that would confirm or disconfirm the proposed cognitive model. However, the success of the training procedure is further converging evidence in support of the proposed cognitive model and is also potentially useful in devising teaching methodologies for other domains.

Chapter 2. Learning Color Words

The Early Word Learning Literature

The early word learning literature attempts to explain the success of toddlers in learning the meaning of words (Woodward & Markman, 1998). Many in the field attribute children's remarkable success to their early understanding of the communicative intentions of others through joint attention (Tomasello, et.al., 2003) or to a set of default assumptions that they use to determine what a word is likely to mean such as the whole object, taxonomic and mutual exclusivity assumptions or constraints (e.g., Echols, 1991; Littschwanger & Markman, 1994; Waxman & Maskow, 1995; Woodward, 1992).

Many psychologists contend that the default assumptions or constraints are not learned but are innate (Soja, Carey & Spelke, 1991; Chomsky, 1959). Others contend that they are pragmatic, that is, learned (Clark, 1987). According to the emergentist coalition model for word learning, these heuristics constitute an expanding tool kit some innate but not absolute and some learned that are used by the child to arrive at the most likely referent for the word being heard (Hirsh-Pasek, Golinkoff & Hollich, 2000).

However, as Nelson has noted:

...children learn the names of things they can act on, whether they are toys, shoes,...as well as things that act themselves such as dogs and cats. They do not learn the names of things in the house or outside that are simply "there"... (Nelson, 1973, p.31).

The first words that children learn are not all nouns and vary considerably from child to child (Nelson, 1973). Nelson's claim is that children learn those words that relate to things that are not simply there, but rather learn those words that refer to objects

and events that are meaningful to the child or to the family and culture in which the child is embedded (Nelson, 2007). The hypothesis of this dissertation is that objects and events are meaningful in Nelson's sense, if they assist the child in understanding his world, that is, if they are not just there, but help her to predict what will happen next. Also consistent with this approach is the evidence that children base concepts on events (McDonough & Mandler, 1998; Nelson, 1985) and function (Nelson, 1974; Kemler-Nelson, 1995) and that early syntactic development emerges from children's understanding of the various scenes that make up their lives (Tomasello & Brooks, 1999).

Nelson (2007) argues that meaning, which after all is what word learning is all about, is significance for the individual. Children and adults want to understand the world and feel stressed in situations in which they have no control and in situations that are unpredictable. What is argued here is that what is significant to an individual are those things, those attributes, those events, those behaviors, those conjunctions, those "whatevers" that impinge upon her sense organs and enable her to predict what is likely to happen next.

The Color Word Literature

As previously noted, the developmental literature has described the learning of color words by children as being both protracted and perplexing as compared to learning nouns, verbs and other property words such as "big" and "little" or "wet" or "dirty" (Sandhoffer & Smith, 1999). The perplexing nature of this phenomenon has been commented upon by such legendary figures as Charles Darwin and Alfred Binet. Indeed, the phenomenon appears to be as robust as children's difficulty with Piaget's

conservation tasks. In fact, the explanation of the two phenomena may be related in that previously learned subconscious heuristics interfere with learning, for example, the toddler repeatedly sees that as liquid is poured into the toddler's cup or glass the level of the liquid rises. The prediction process is engaged and the child learns the heuristic that the higher the level of the liquid in the container, the more liquid there is, which is a perfectly reasonable conclusion that is more often right than not, in the toddler's world. To learn that this is not always the case, requires not only the appropriate experience but also the ability to inhibit the previously learned heuristic (the Inhibition Process).

While some toddlers younger than two know that color words such as blue, red and green are appropriate answers to questions asking about colors as, for example, "what color is this banana?", the toddler is as likely to answer blue as yellow. Until about age four or five, some children continue to have difficulty in accurately verbally naming the color of objects (hereinafter "color naming") long after they are able to accurately identify other perceptual attributes of objects such as shape (Bornstein 1985). This robust developmental anomaly was noted by Charles Darwin. He initially concluded that his own children were born color blind since they were seemingly unable to learn to accurately name colors despite repeated explicit instruction. As quoted in Bornstein (1985, p.388), Darwin wrote:

I carefully followed the mental development of my small children, and I was astonished to observe in two or, as I rather think, three of these children, soon after they had reached the age in which they knew the names of all the ordinary things, that they appeared to be entirely incapable of giving the right names to the colours of a colour etching. They could not name the colours, although I tried repeatedly to teach them the names of the colours. I remember quite clearly to have stated that they are colour blind. But

afterwards this turned out to be an ungrounded apprehension. When I told this fact to another person, he told me that he had observed a rather similar case. The difficulty which small children feel, whether in discrimination or, much more probably, in naming the colours, seems therefore to merit further investigation.

Yet, though preschool children have difficulty in color naming, Alfred Binet, as quoted in Bornstein (1985, p.391), at the turn of the century wrote:

The young child distinguishes, recognizes, and easily matches without the least hesitation the most delicate shades of color, and has nothing to envy in the adult so far as his color sense is concerned; it is the verbalization of his color sense, if we may so express it, in which he is defective. Bornstein's 1985 translation.

Indeed, subsequent research has definitely established that infants as young as three months are able to discriminate all four primary colors and even neonates appear to be able to distinguish red from gray (Slater and Johnson 1998). Thus, the difficulty that preschool children have in color naming does not stem from an inability to distinguish colors because of an immature, poorly functioning color perceptual system. Moreover, the fact that toddlers reply to questions of the form "what color is this?" with a color word albeit not necessarily the correct color and are also able to discriminate and match colors would seem to imply that they recognize color as a "category" and thus would seem to possess a "concept" of color (Soja, 1994; Rice, 1980).

Darwin's observation that repeated instruction is ineffective in teaching young children to accurately label colors was convincingly demonstrated by Rice (Rice, 1980). Rice designed and implemented a series of experiments to determine whether for preschool children pre-lexical conceptual organization of a domain was either necessary, facilitated or was irrelevant to the ability to master the use of words that correspond to the

conceptual domain. The domain she explored was color and her ten participants ranged in age from 2;3 to 3;4. She divided her subjects into two groups, color concept users and non-color concept users. The color concept users group was composed of children who were able to match objects by color in a sorting task.

The training procedure for teaching color names used by Rice was for the experimenter to pick up or point to an object and ask “What color is this?” Correct responses were socially reinforced (except initially a food reinforcer was given to the youngest children). If the child’s answer was incorrect, corrective feedback was given to the child by the experimenter such as “No, that’s a green one,” and the item was immediately repeated.

Children were trained to discriminate first between red and green and then red, green and yellow. The results were that the mean number of trials to success – where success was defined to be 80% correct on untrained object probes, across two sessions – was 693.3 for color concept users as a group and 1,338.7 for non-color concept users as a group.¹

A variety of explanations have been put forth to explain this pattern of development. Bornstein (1985) compared the ability of 3-year-olds (3;0 - 3;11) to learn color words (blue, green and yellow) versus shape words (square, triangle and circle). The shapes were all solid black and the colors were represented by blue, green and yellow circles. Citing Wertheimer (1938) and Becker & Kay (1969), Bornstein argued that both the shapes and the colors were simple example stimuli from each of their

respective dimensions. The children were asked to learn an association between each of the colors and shapes and a child's name after being told that, for example, the square is Bill's favorite shape. In his study, 94% of the children learned the shape-name association as opposed to 69% for the color-name association even after 36 trials; the children learned the shape-name association significantly faster than the color-name association; and all the children committed significantly more errors when learning the color-name association as opposed to the shape-name association.

Bornstein concluded that the results of his study eliminated relative perceptual salience, differential experience and individual differences in cognitive development as explanations for the difficulty children have had with color naming. He speculated that perhaps Darwin was partially correct in that while his children were not color blind, the explanation for the difficulty they had in color naming may have had a neurological basis after all. He based his speculation on the work done with respect to one of the subgroups of color dysnomics for whom there is some evidence to believe that they suffer from a disconnection between visual and language cortical structures. (See Oxbury, Oxbury & Humphrey, 1969).

However, Rice's work shows that two year olds can be taught to correctly identify colors, albeit with great difficulty. Moreover, such a specific immature connection between visual cortical structures and language structures affecting the ability to correctly identify color, but not the ability to correctly identify shape while possible seems unlikely.

Nancy Soja (1994) showed that children who cannot correctly identify colors nevertheless have the conceptual foundation necessary to use colors to solve similarity problems (provided the objects are identical except for color) and even to learn different proper names assigned to toy animals where the only difference between the toy animals was their color. This is consistent with Rice's findings that the group of children who were color concept users nevertheless had great difficulty in color naming.

Soja argued that the difference between learning to correctly name colors and being able to succeed in the proper name mapping task is that the latter task is not one of reference but rather of co-occurrence. The proper name co-occurs with the color. However, the name refers not to the color, but rather to the particular toy animal in contrast to the identical toy animal of a different color. The different colors establish identity, that is, the children are able to learn that the red tiger is "Tony" and the green tiger is "Jennifer" even though they cannot correctly name the color of either!

Since the children in Soja's study were able to discriminate among colors and were able to use colors to solve identity tasks, Soja concluded that the difficulty encountered by children in learning to correctly name colors with color words was because of word learning constraints² – perhaps more aptly described as subconscious word learning heuristics – that militate against learning color words.

One such heuristic is that children come to expect or predict that adjectives refer to relative binary contrasts such as big/little, slow/fast, wet/dry, clean/dirty and so on (where the very meaning of each adjective in the contrast requires an understanding of

both). (Landau and Gleitman, 1985). If that is the case, then color words would violate that expectation because there are many colors and color terms that are absolute, that is, blue is blue and it is not blue in contrast to yellow or green. The fact that color words also violate the hypothesized heuristics of contrast (Clark, 1987) and mutual exclusivity (Markman and Wachtel, 1988) could, Soja argued, also contribute to the difficulty children have in learning to use color words correctly. The contrast heuristic (which, it should be noted, is often violated), states that children assume or expect that two different words do not have the same meaning. If true, then the color domain would be difficult for children to learn since many different words are used to name colors that are quite similar to one another such as blue and navy. (However, common observation would suggest that when first teaching colors parents do not confuse their children by trying to teach them to distinguish between blue and navy at a time when they are having difficulty with correctly identifying blue versus red.) The heuristic of mutual exclusivity is the converse of the contrast heuristic, namely a single reference is expected by the child to be labeled by one word.

Several investigators (Au and Markman, 1985; Macarrio, 1991) have suggested that children tend to assign words to properties that co-occur or predict other properties and color rarely identifies an object's function or any other property of the object. Soja noted that while this might not at first blush appear to be a plausible explanation for the difficulty children have in learning color words in light of the fact that some of the early adjectives that children learn describe a state and sometimes a temporary state at that, such as "wet," "broken," and "dirty," these states are the result of and can participate in

causal interactions, that is, it makes a difference if something is wet as opposed to dry. However, colors, at least in the ordinary experiences of children, do not change as a result of causal interactions, nor do colors as colors cause anything to happen, that is, the fact that the ball is red or green makes no difference in how the ball will behave when thrown.

Sandhoffer and Smith (1999) studied and compared the developmental trajectory of learning color words and size words.³ The subjects ranged in age from 1;8 to 2;9. Except for one child, the children first learned word-word maps, that is, they answered the question “what color is this?” with a color word. Second, the children learned word-property maps, that is, they were able to comprehend and could succeed at the task “show me the red one.” Third, they learned to answer the question “what color is this?” However, although the children could tell you that the cat was red and the cup was red, they still had difficulty in the fourth and final task of making property to property maps as tested by the question “what matches this?”

Soja’s (1994) findings were seemingly contrary (and consistent with the Bearison & Siegel findings (Bearison & Siegel, 1968)) namely, that toddlers could match objects by color even though they could not correctly name the color of objects. Sandhoffer and Smith argued that Soja’s matching task was structured so that none of the objects – red suitcase, black horse and red disk – were similar on any dimension other than color and was, therefore, solvable by using overall similarity while their task was not. However, one of Sandhoffer and Smith’s tasks was similar to Soja’s task in that the only dimension of the object to be matched that was similar was color and while their other task was

more difficult in that some of the objects matched on a dimension other than color, the performance on the two tasks did not differ and were combined by them for analysis. Nevertheless, it is possible that Soja's contrary findings may be explainable by the fact that the Sandhoffer and Smith task was more complex in that it required the child to match the target object by choosing one object from an array of six some of which matched on a dimension other than color. Likewise, the color form preference literature (Bearison & Siegel, 1968) which also appears contradictory to Sandhoffer and Smith's findings can be distinguished because most of the work in this literature involved the use of a forced choice categorization task as opposed to a matching task with simple shapes such as triangles, circles and squares rather than complex shapes and the difference in tasks may explain the difference in results.

Sandhoffer & Smith noted that children's ability to match objects by color lagged behind their comprehension of color words in their study seemed counterintuitive. Moreover, their study showed that this perplexing pattern of acquisition did not apply to size words. As a possible explanation, they pointed to the work of Gentner and Ratterman (1991) and Kotovsky and Gentner (1997). Those studies produced evidence supporting the hypothesis that training in explicitly comparing objects promotes selective attention to the property in question allowing one to abstract the designated property. Those studies also showed that learning names for comparisons facilitates understanding of the conceptual basis underlying such comparisons. Sandhoffer and Smith speculated that explicit training in making property to property maps is required before children will be able to succeed at this task. They also suggested that learning color labels may help

children in attending and, therefore, abstracting the property relevant to succeed at the matching task.

Note that these studies and speculations are consistent with the Prediction Process, Filtering Process and Inhibition Process hypotheses in the following ways. Children are not attending to colors because they learn to ignore colors if colors do not predict anything, that is, as Nelson put it, if they are just there (the Filtering Process). Yet Soja found that children could learn the names assigned to two toys that were identical in all respects except for color which shows they can and do attend to the color of objects when color makes a difference (the Prediction Process). Finally, the difficulty in color naming may, as Soja claimed, arise partly because color names violate word learning heuristics that children have learned by virtue of the Prediction Process and have difficulty in inhibiting these word-learning heuristics in circumstances in which they do not work (the Inhibition Process).

Interestingly, Sandhoffer and Smith did not comment on Soja's far more perplexing result, namely, that toddlers could easily learn to associate colors with proper names before they were successful at color naming. A possible explanation of this surprising result is that color does have a function in the child's world—it can be used to identify (predict) to whom objects belong. So, for example, as Soja noted, the child learns that the red tooth brush belongs to mother and the green toothbrush belongs to the child. Thus, the child learns early on that color can be used to predict identity or ownership. However, even the identity heuristic works against learning correct color

identification, because the fact that mother's toothbrush is red does not mean that the red car is hers and, in fact, the red car may belong to father.

One explanation that has been offered that is consistent with Nelson's view (Nelson, 1973) and with the Prediction Process, Filtering Process and Inhibition Process is that children have difficulty in learning the "meaning" of color, because in the everyday life of the very young child color by itself is not predictive of function or anything else that is meaningful or useful to the child (Au and Markman, 1987).

Since the color of an object is generally not predictive of what comes next, children would over time learn not to pay attention to color just as they learn not to pay attention to phonetic contrasts that do not make a difference in their native language (Maye, et.al., 2002). This learned inattention taken together with the fact that color words do, as Soja pointed out, violate many of the predictive word learning heuristics that children do learn, would seem sufficient to account for the difficulty children have in learning to use color words. This suggests that children who are taught colors using a training regimen in which the colors are predictive of an outcome may more quickly learn the "meaning" of color and will succeed at color naming tasks if, and only if, they can overcome the learned heuristic that color is only a marker for identity or ownership and have the ability to inhibit previously acquired word learning heuristics that conspire together to interfere with learning.

Chapter 3. Experiment I – Learning Colors

Research Questions

While the literature and common experience show that children in the age range of two through three have little difficulty in learning the names of common shapes, the literature from Darwin to Rice just as clearly shows that the usual method of trying to teach color words to young children by asking them “what color is this?” and then providing positive or negative feedback, as applicable, does not work very well, if at all. If, besides violating the common word learning heuristics, the difficulty children have in learning to correctly use color words stems from learned inattention resulting from the fact that color is not in general predictive of function or use or anything else of consequence except perhaps in the limited cases of identity and ownership, then it is not surprising that the usual technique used by Darwin, Rice and others to teach color naming is ineffective. If the problem is learned inattention and conflict with word learning heuristics, then the teaching methodology must be one in which the color of objects engages the Prediction Process.

The specific research question addressed in Experiment I described below is:

Will a teaching method that engages the Prediction Process cause children to attend to the color of objects and more quickly learn to correctly identify the color of objects than children taught with a teaching method that does not engage the Prediction Process?

General Description

Children were given a pre-test to determine whether they could correctly use color words. Those who could not were divided into an experimental group and a control group. Each participant in both the experimental group and control group were individually taught the colors blue, red and yellow using the same materials. Both groups were taught using a board on which twelve mittens, twelve cars and twelve flowers made out of cardboard cutouts were fastened in no particular order. Four of each of the mittens, cars and flowers were blue, four were red and four were yellow.

As will be described in more detail below, the experimental and control teaching procedures in addition to using the same materials, had the same number of sessions and trials within each session, provided the same feedback and were otherwise identical in all respects except that the experimental teaching procedure was designed to enable the child to predict the color of the object the investigator would point to next, thereby engaging the Prediction Process, while the control training procedure did not enable the child to predict what color object the investigator would point to next. Specifically, in the experimental training procedure, the investigator touched five of the twelve red objects consecutively saying “This is a red mitten, this is a red car, this is a red flower” and so on for the first five objects. The investigator then consecutively touched the remaining seven red objects asking “This is a _____?” In the control training procedure the first five objects touched were not all the same color. Thus, the investigator may have touched a red car, blue mitten, yellow flower and so on. The investigator would then touch seven more objects not all having the same color asking the same question as was

asked in the experimental training procedure. In both cases, feedback was given to the child.

Participants:

The participants were either children attending a pre-school head start program or its day care program in a medium sized city in Westchester County, New York. Almost all of the participants were fluent in both English and Spanish. However, for the three or four participants who were more comfortable in Spanish, the questions were posed to them in Spanish and answers in Spanish were accepted.

A total of one hundred sixty one parental consents were obtained. Eighty four for the 2009 Spring semester and 77 for the 2009 Fall semester. A total of 33 of the 161 were not able to correctly identify all three colors as determined by the pre-test described below. Of those 33, two experimental participants were dropped, one being a special needs child and the other being uncooperative and four control group participants were dropped, two because they moved, one because he was a special needs child and the other because he was uncooperative. For each semester the participants were arranged in birth order, a coin was tossed to determine whether the youngest would be assigned to the experimental or control group – in the Spring semester the youngest was assigned to the control group and in the Fall semester the youngest was assigned to the experimental group – and the remaining participants were then alternately assigned to the experimental group or control group, as the case may be. The participants ranged in age from 3;1 to 4;9 in the control group with a mean age of 3;11 and from 2;10 to 4;11 in the experimental group with a mean age of 4;1. Five of the twelve participants in the control

group were females and six of the fifteen participants in the experimental group were females.

Materials:

Twelve mittens, twelve cars and twelve flowers were attached by Velcro to a board so that their positions could be changed (hereinafter “training objects”). The training objects in each set were made of construction paper covered with clear plastic, were approximately four inches by two inches and identical in all respects except for color. Four of each set of the training objects were blue, four were red and four were yellow. These colors were chosen as they are common and red and blue were the colors used in the previously cited studies coupled either with green or yellow.

The pre-test was designed based on previous research particularly that of Sandhoffer & Smith (1999) which generally showed that a child’s success on any one of the tasks tested in the pre-test described below does not necessarily imply that the child will be successful on all of the other tasks.

Procedure:

Pre-Test:

Each of the children for whom consent was obtained was tested as follows:

(1) First the investigator handed to the child one at a time three identical rubber octagons one red, one blue and one yellow each approximately $1\frac{7}{8}$ inches by $1\frac{7}{8}$ inches and asked the child to answer the question “What color is this?” If the child responded to all three correctly, they were excluded from the study;

(2) The investigator then placed in front of the child an array of three shapes – either three circles, three squares or three triangles – made of construction paper and approximately 2½ inches by 2½ inches and identical in all respects except that one was red, one was yellow and one was blue. The investigator then requested the child to “Show me the red one”. The investigator noted the child’s answer and the correct answer. No feedback was given to the child. This procedure was repeated with new object arrays with the investigator asking the child to “Show me the yellow one” and then “Show me the blue one”;

(3) The investigator then placed in front of the child three identical trapezoids made of construction paper and approximately 3 inches by 2 inches one red, one blue and one yellow. The investigator pointed to each of the objects in turn asking the child the question: “What color is this?” The investigator noted the child’s answer and the correct answer. This test was the same as the first test other than that different objects were used and gave the child a second chance to show his or her knowledge after having heard the name of the colors in the immediately previous “Show me” section of the test. No feedback was given to the child; and

(4) The investigator then placed in front of the child an array of a circle, rectangle and triangle made of construction paper each approximately 2 inches by 2 inches each of a different color and a target shape to be matched by color – either a triangle, rectangle or circle made of either red, blue or yellow construction paper and of a similar size. While pointing to a shape to be matched by color, the investigator then asked the child to “Show me the one that is the same color” from an adjacent array of

three shapes which contained the same shape as the target shape to be matched but of a different color and a different shape having the same color and a third shape that was different on both dimensions. The target color to be matched might be a red circle and the array a red square, a yellow triangle and a blue circle. The investigator noted the child's answer and the correct answer. No feedback was given to the child.

Except for the first task, one point was awarded for each question answered correctly by the child.

Experimental Training Procedure:

The training sessions were conducted over a period of several weeks outside of the classroom but either in the school kitchen or school office as the investigator was not permitted to be alone with a participant. Not counting the time involved in taking and returning a participant to the classroom, the training procedure took between five and ten minutes. The investigator touched five of the red objects, one after the other stating as each red object was touched that "this is a red [name of training object]." The investigator then touched another of the remaining seven red training objects and asked "This is a ____?" and noted the participant's answer. The investigator did give the child feedback and then touched the next red training object and again asked the child "This is a ____?" Once the investigator had touched all the red training objects, the entire procedure was repeated for the blue and yellow training objects. The training procedure was repeated on subsequent training sessions. At the end of each training session each participant was given one sticker.

Control Training Procedure:

The training procedure and conditions for the control group were identical to that for the experimental group except that the investigator was as likely to touch a yellow, blue or red training object after touching a red training object. Thus, the child was not able to guess the color of the training object based upon the color of the training object last touched by the investigator.

Post-Test:

After ten training sessions were completed (or earlier, if the participant correctly answered all of the investigator's queries during the training sessions) each of the participants in the control group and experimental group was given a Post-Test identical to the Pre-Test.

Results:

The experimental training procedure was designed to engage the hypothesized Prediction Process which is posited to be the natural learning process. As a consequence, the expectation was that the participants who were taught colors using the experimental training procedure would learn to correctly verbally identify the color of objects more quickly than the control group who were taught using the procedure used by Rice and the other investigators cited above.

Table 1 below sets forth the mean scores and standard deviations for each subtest of the Pre-Test as well as the total Pre-Test Scores for the fifteen experimental group participants and the twelve control group participants.

Table 1
Pre-Test Results

	Experimental N=15		Control N=12	
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Mean</u>	<u>Standard Deviation</u>
“Show Me” Subtest	1.67 ^a	0.79	1.67 ^a	0.83
“What Color?” Subtest	0.67 ^a	0.70	0.75 ^a	0.83
“Matching” Subtest	2.0 ^a	1.03	1.67 ^a	1.38
Total Test Score	4.33 ^b	1.74	4.08 ^b	1.37
^a 3 possible				
^b 9 possible				

There were, as expected, no significant differences between the means of the experimental group and control group on any of the Pre-Test subtests nor was there a significant difference in the means of the combined total scores achieved by each group. However, as Table 2 shows below, the mean scores of the participants on the “Show Me” and “Matching” subtests were significantly higher than the mean score on the “What Color?” pre-test while there was no significant difference between the mean scores on the “Show Me” and “Matching” subtests.

Table 2 sets forth the means and standard deviations for each of the subtests combined for both the experimental and control group as well as significance levels of the differences in the means between the “What Color?” subtest and the “Show Me” subtest and the “What Color?” subtest and the “Matching” subtest and between the “Matching” subtest and the “Show Me” subtest.

Table 2
Pre-Test Results Combined for Both Groups

	Show Me N=27	What Color? N=27	<u>t</u>	<u>p</u>
Mean (3 possible)	1.67	0.7	4.39	<.0001
Standard Deviation	0.86	0.76		
<hr/>				
	Matching N=27	What Color? N=27	<u>t</u>	<u>p</u>
Mean (3 possible)	1.85	0.7	4.26	<.0001
Standard Deviation	1.18	0.76		
<hr/>				
	Show Me N=27	Matching N=27	<u>t</u>	<u>p</u>
Mean (3 possible)	1.67	1.85	-.064	>.05
Standard Deviation	0.86	1.18		

Set forth in Table 3 below are the mean scores and standard deviations for both the experimental group and control group for each sub-test of the post-test and the total post-test scores as well as the significance level of the difference in the means of the two groups.

Table 3
Post-Test Results

	Experimental N=15		Control N=12		<u>t</u>	<u>p</u>
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Mean</u>	<u>Standard Deviation</u>		
“Show Me” Subtest	2.67 ^a	0.7	2.33 ^a	1.31	0.811	.219
“What Color?” Subtest	2.6 ^a	0.71	1.92 ^a	1.03	1.946	.047
“Matching” Subtest	2.87 ^a	0.50	2.25 ^a	1.12	1.781	.052
Total Test Score	8.13 ^b	1.26	6.5 ^b	2.73	1.912	.049

^a3 possible

^b9 possible

Except for the “Show Me” subtest and the matching subtest where the difference was marginally significant ($p > .05$), the difference in the post-test mean scores for the experimental group versus the control group with respect to the key “What Color?” subtest scores ($p = .04$) and the total post-test scores ($p = .04$), were significant.

Table 4 below shows the mean scores and standard deviations as well as the significance levels of the difference between the mean scores for the experimental and control groups with respect to the increase in the total scores achieved from pre-test to post-test, the increase in the combined total scores from pre-test to post-test for the “Show Me” and “What Color?” subtests and for the “What Color?” subtest by itself.

Table 4
Post-Test Minus Pre-Test Results

	Experimental N=15		Control N=12		<u>t</u>	<u>p</u>
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Mean</u>	<u>Standard Deviation</u>		
Increase from Pre-Test to Post-Test in Total Scores	3.8	2.78	2.42	2.09	1.637	.07
Increase from Pre-Test to Post-Test in “Show Me” and “What Color?” Subtests Combined	2.93	1.29	1.83	1.49	2.054	.03
Increase from Pre-Test to Post-Test in “What Color?” Subtest	1.93	0.77	1.17	1.21	1.891	.04

Except for the increase in total scores which was marginally significant ($p=.07$), the increase for scores in the “Show Me” and “What Color?” subtests combined, (that is, after excluding the marginally relevant “Matching” subtest scores) was significant ($p=.03$) as was the increase in the key “What Color?” subtest score ($p=.04$).

Table 5 shows the mean increase in total scores from the pre-test to the post-test and the standard deviation for the six participants in the experimental group who were under four years of age and for the six participants in the control group who were under four years of age. The mean score increase of this subset of the experimental group on all subtests was 4.67 versus a mean score increase of 2.83 for this subset of the control group and 0.83 versus 1.83 increase for the “What Color?” subtest. Although the effect size was relatively large in both instances, the significance level $p=.089$ for total scores and

P=.073 for the “What Color?” subtest was not significant in each case because of the extremely small sample size.

Table 5
Post-Test Versus Pre-Test Results
Comparison of Mean Increase in Total Scores
for Participants Under Four Years of Age

	Experimental N=6		Control N=6		<u>t</u>	<u>p</u>
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Mean</u>	<u>Standard Deviation</u>		
Increase from Pre-Test to Post-Test in Total Scores	4.67	1.66	2.83	1.95	1.756	.073
Increase from Pre-Test to Post-Test in “What Color?” Subtest Scores	1.83	0.90	0.83	1.16	1.67	.082

Table 6 shows the mean increase in total scores and the mean increase in the “What Color?” subtest scores from pre-test to post-test and the standard deviation for the eight participants in the experimental group and the eight participants in the control group who had a total score of 4.0 or lower on the pre-test, that is, a total score below the mean total score for each group as a whole.

Table 6
Comparison of Mean Increase in Total Scores
for Participants Scoring Below the Mean
Total Score on the Pre-Test

	Experimental N=8		Control N=8		<u>t</u>	<u>P</u>
	<u>Mean</u>	<u>Standard Deviation</u>	<u>Mean</u>	<u>Standard Deviation</u>		
Increase from Pre-Test to Post-Test in Total Scores	5.0	1.66	2.63	2.18	2.446	<.03
Increase from Pre-Test to Post-Test in "What Color?" Subtest Scores	2.0	0.87	0.75	1.2	2.385	<.03

Finally, Table 7 compares the proportion of (i) the eight participants in the experimental group whose combined pre-test subtest scores were below the mean score of 4.33 achieved by the experimental group as a whole and who learned at least two of the three colors, namely, 7 out of 8, with (ii) the 3 out of 8 proportion of the eight participants in the control group who learned at least two out of the three colors and whose combined subtest scores were below the mean score of 4.08 of the control group as a whole. The difference in the proportions even for this small sample size using Students distribution for seven degrees of freedom was significant at $p=.041$.

Chapter 4 Experiment II – Learning Shapes

Method

Research Questions:

A second study was done using the same two teaching methodologies to teach the names of shapes that the participants did not know. The purposes of the study was to determine whether learning the names of shapes was as the literature reported easier than learning to name the color of a shape and whether the experimental training method was more effective than the control training method in teaching the name of shapes.

Participants:

The participants were drawn from the same head start program. There were thirteen participants ranging in age from 4;0 to 4;10 with a mean age of 4;4. Of the seven participants in the experimental group, three were female and the mean age was 4;5. Of the six participants in the control group, four were female and the mean age was 4;4.

Materials:

Three shapes were used – a trapezoid, pentagon and octagon made of black poster board and approximately 2 inches by 2 inches. Twelve of each type of shape were fastened to a board in no particular order.

Pre-Test:

The investigator pointed to each shape in turn and asked the child to name the shape. No feedback was given.

Experimental Training Procedure:

The experimental training procedure and the conditions were identical to that used in Experiment I except that the investigator simply asked “What shape is this _____?” as opposed to “What color is this [name of training object]?”

Control Training Procedure:

The control training procedure and conditions were identical to that used in Experiment I with the same exception as noted above with respect to the experimental training procedure.

Post-Test:

The Post-Test was administered after 10 training sessions were completed (or earlier, if the participant answered all the investigator’s questions correctly during the training session). The Post-Test was identical to the Pre-Test.

Results:

Table 7 shows the comparison of the mean score increase in the shape naming experiment for both control and experimental groups combined (13 participants) from pre-test to post-test in comparison to the mean score increase in the “What Color?” subtest for both control and experimental groups combined (27 participants) and the significance level of the difference between the means.

Table 7
Comparison of Increase in Mean Scores
for Colors Versus Shapes

	Increase in "What Color?" Scores <u>N=27</u>	Increase in "What Shape" Scores <u>N=13</u>	<u>t</u>	<u>p</u>
Mean	1.59	2.69	4.797	<.0001
Standard Deviation	0.99	0.46		

Chapter 5. Discussion

This research is based on the hypothesis that the primary learning process used by the brain is to predict what is going to happen next in the world and then comparing its prediction to what in fact happens next. Coupling this hypothesis with the fact that the brain has limited processing power, leads to the corollary hypothesis that those aspects of a child's environment that do not enable a child to predict what is going to happen next or otherwise do not assist the child to understand her environment or participate in activities with others are not attended to by the brain. While a variety of explanations have been offered to explain why children have difficulty in learning to correctly identify the color of objects from immature connections between the visual cortex and the language center (Bornstein, 1985) to word learning constraints (Soja, 1994), the explanation offered here is that the brain of a young child has learned to ignore (that is, it has learned to filter out as not important) the color of objects because in its limited experience colors do not predict what is going to happen next.

In order to test this hypothesis, a teaching method was devised that enabled the child to guess or predict the color of the object the investigator would point to next in order to engage the Prediction Process and thereby cause the child to attend to the color of the object. In fact, in both the color and shape studies, the children in the experimental group understood by the second training session, if not during the first training session, the "follow the leader" or "guessing game" nature of the training procedure. They generally provided the correct answer with respect to the color or shape of the remaining training objects the investigator pointed to after the investigator had named the first five

training objects in the series. Indeed, many of the children interrupted the investigator after the first few sessions and began correctly naming the color or shape of the training objects before the investigator had completed naming the first five exemplars. The children were, as hoped, guessing which training object would be pointed to next. The Prediction Process appears to have been engaged.

Because of the difficulty in obtaining cooperation from pre-schools and/or parents of children in the age range of previous studies, the majority of the participants in this study were four years of age or older and of the twenty seven participants only one was under three (2:10) and only five were under three and a half years of age. Nevertheless, in confirmation of the literature, even in this somewhat older age range approximately 20% of the children in the pre-school head start population for which permission was obtained still had difficulty correctly using color words. Yet Experiment II established that children from this same population and age range had little difficulty in learning to name shapes no matter which training regime was used. This result was again consistent with the literature and common experience.

Experiment I

Experiment I confirmed that the experimental training method designed to engage the Prediction Process was more effective than the control training procedure in teaching children to correctly identify the color of objects ($p=.049$ for the “What Color?” subtest). The effect size, that is, the difference in the increase in the mean number of correct answers from pre-test to post-test being 1.93 for the “What Color?” subtest out of a total

possible increase of 2.33 (83%) for the experimental group and an increase of 1.17 for the control group out of a total possible increase of 2.25 (52%).

However, since (1) those participants in the study (average age of 4;0) were older than those in the Rice, Sandhoffer & Smith, Soja and Bornstein studies cited earlier (ages two through three), (2) the majority of the participants were over four years old and ten four and a half years old or older and (3) the participants were simultaneously being taught colors in their pre-school classes, one would expect that a number of the participants in both the control and experimental groups would learn to identify the color of objects over the three month course of the study without the additional instruction provided by the Investigator. All of these factors would, if anything, tend to reduce the difference in the increase in the means between the two groups because of a ceiling effect.

Thus, it is instructive to also look both at those participants who were under four years of age and those participants whose aggregate subtest scores were below the mean total pre-test scores of their respective groups (4.33 for the experimental group and 4.08 for the control group). Those comparisons show that the increase in the mean score for the "What Color?" subtest from pre-test to post-test for the six experimental participants versus the six control participants who were under four years of age was 1.0 ($p=.082$) as compared to the effect size of 0.76 ($p=.04$) for the two groups as a whole. Moreover, if the results from the pilot study with two children both under three years old in age from a middle class day care center who both went from knowing no colors to all three colors

are combined with the six experimental group participants under four years of age, the result would have been significant ($p=.044$, $t=2.55$) as opposed to $p=.082$.

Likewise, if we compare the aggregate increase in scores from pre-test to post-test for the eight experimental participants and the eight control participants whose Pre-Test scores were below the mean pre-test scores, the effect size was 2.37 ($p<.03$) in comparison to an effect size of 1.38 ($p=.06$) for the two groups as a whole. The fact, that in one case the difference in the means was marginally significant and in the other case significant, given the small sample sizes, is noteworthy, and further suggests that a ceiling effect was depressing the difference between the two groups as a whole.

The analysis is further strengthened by comparing the seven out of eight proportion of the experimental group who learned at least two of the three colors versus the three out of eight proportion of the control group participants who went on to learn at least two out of the three colors. Even for this small group size the difference was significant at $p=.041$.

The results also confirmed that children learn to match by color (Bearison and Siegel, 1968; Soja, 1994) before they learn to correctly name the color of objects ($p<.0001$) and are able to show you the red object before they can correctly name the object as being red ($p<.0001$) (Sandhoffer and Smith, 1999). However, these results do not support the Sandhoffer & Smith claim that the "Matching" task is learned after the "What Color?" task. Instead, the results reported in Table 2 ($p<.0001$) are consistent with Soja's contrary finding. It should be noted that while the matching test used in this study

was more complex than Soja's task it was less complex than the one used by Sandhoffer & Smith as described earlier.

Sandhoffer and Smith (1999) emphasized the role of selective attention in learning color words and other words as have other investigators. See Au & Markman (1965) and Macario (1991). Yet in order to succeed at the "Show Me" task, children must be able to attend to the color of the object. While on the surface then it appears quite perplexing that children can succeed at the task "Show Me the red one" long before they are able to successfully name the object as being red, I believe the explanation of this counter-intuitive result is that the Prediction Process first naturally leads to children attending to colors as only a marker of identity as Soja demonstrated (1994). "Show Me the red one" is in a sense using color as a marker of identity.

Experiment II

Experiment II confirmed the findings in the literature regarding the relative ease with which children between the ages of two and three learn the names of common shapes as compared to learning to correctly use common color words and extended the findings to four year olds and novel shapes.

Specifically, children learn the names of shapes and to correctly apply them much more easily than they learn to correctly name the color of objects ($p < .0001$). This was true even though the names of the colors were common – red, blue and yellow – and the names of the shapes were unfamiliar – trapezoid, pentagon and octagon – and the shapes were not as distinguishable as those used by Bornstein (1985) in his study of three year olds – circle, square and triangle.

Although in the shape naming study the experimental group did marginally better than the control group – five out of seven versus four out of six learned to correctly identify all three shapes, the difference was not significant. However, there was clearly a ceiling effect as all of the children learned to correctly identify at least two of the three shapes and all but four out of thirteen learned to correctly identify all three shapes. Moreover, there was a design flaw in the experimental training procedure for the shape experiment which resulted in some of the participants in the experimental group not paying attention to the shape the investigator was pointing to which would have had an adverse affect on the learning experience for those participants in the experimental group. In both Experiment I and Experiment II the experimental group immediately understood how the “game” should be played, that is, if the investigator started off saying red or trapezoid, then all the following objects that would be pointed to would either be red or a trapezoid until the investigator turned to a new color or shape and started over again. In Experiment I, in order to give the right answer, the participant not only had to say the correct color but also whether it was, for example, a red mitten, car or flower. This required the participant to pay attention to the training objects pointed to by the investigator. However, this was not the case for the shape study as all trapezoids were identical in all respects. Thus, some of the participants in the shape study simply answered trapezoid for the remaining items without attending to the object the investigator was pointing at. While this problem was discovered early in the training, it was decided to complete the experiment to see if learning the names of unusual shapes

for four year olds was easier than learning to correctly identify objects by color as, in fact turned out to be the case.¹

Conclusions and Implications

The focus of this study was to determine whether the experimental training procedure which engaged the Prediction Process was more effective in enabling pre-school children to learn to correctly identify the color of objects than was the control training procedure which was the training procedure used in the cited studies as well as the procedure commonly used by parents and teachers to teach color words.

In summary, the results demonstrated that the experimental training procedure which engaged the Prediction Process was more effective than the control training procedure based upon (1) comparing the mean score on the post-test in the key “What Color?” subtest, ($p=.047$) as well as the increase in the mean scores from pre-test to post-test for the “What Color?” subtest ($p=.04$) and (2) comparing the means of the total post-test scores ($p=.049$). However, the results are limited by the relatively small number of participants in the study, by the fact that they were simultaneously being taught to correctly identify the color of objects in their pre-school class and because many of the participants were over four years of age, an age at which they would ordinarily learn to correctly identify the color of objects in any event. However, as noted earlier, the last two factors would be expected to make it more difficult to establish a significant difference between the experimental and control groups because of a ceiling effect.

The success of the experimental training procedure suggests that teaching material to young children that does not naturally engage the Prediction Process such as learning to recognize the letters of the alphabet and learning to pronounce or recognize written words would benefit from using a teaching methodology that engaged the Prediction Process. So, for example, instead of teaching words by using flash cards with single words, cards with four or so words such as top, mop, hop and pop, printed multiple times in random order with the words being taught using the experimental training procedure described in Experiment I may be more effective. To test the effectiveness of such a teaching methodology would be a simple extension of the research described here.

Similarly, an abstract concept such as Newton's first law of motion which is counter-intuitive because in the real world friction and gravity prevent objects from moving in a straight line with constant velocity could be more effectively taught through computer simulations that engage the Prediction Process. So, for example, a computer simulation could be designed to show a series of different objects propelled and set free to traverse different surfaces such that as the objects and surfaces become obviously smoother, the object clearly travels further and further in a straight line so that it becomes intuitively obvious that absent friction an object would continue indefinitely in a straight line at a constant velocity. In short, the success of Experiment I suggest further research in designing and using teaching methods that engage the Prediction Process may prove fruitful in domains other than color naming.

The implications of the Prediction Process for developmental psychology in general is that while researchers often use the language that the child expects or

anticipates, they often do not explicitly consider that the child is actually predicting what is going to happen next and learning by comparing the prediction the child makes with what, in fact, does happen next. Thus, before a researcher gives up, so to speak, and posits a nativist “explanation” for the acquisition of knowledge or heuristics in a particular domain, the researcher should consider whether the stimuli that the infant or young child encounters on a daily basis can together with the Prediction Process, Filtering Process and Inhibition Process explain the acquisition of knowledge or heuristics in that domain including, as Tversky and Kahneman found, the acquisition of incomplete knowledge or misleading heuristics. So, for example, one of the arguments in favor of Chomsky’s “language organ” is that the child does not receive sufficient negative input to enable the child to learn grammar. This argument falls away if the child is generating her own negative evidence by virtue of the Prediction Process.

This research constitutes corroborating evidence for the Prediction Process in that it makes sense of a developmental anomaly noted first by Darwin and confirmed by subsequent research including the counter-intuitive finding within this research that children can successfully succeed at the “Show Me” the red object task (because color is predictive of ownership or identity), but still not consistently be able to correctly identify the color of red objects on the “What Color?” task.

The Prediction Process is consistent with most developmental theories and/or the findings motivating such theories including neural network modeling. If one also posits that an heuristic once learned even if incorrect is either never forgotten or only slowly forgotten and that development occurs not by modifying such previously learned

heuristics even if they are erroneous or only applicable in certain limited circumstances but by learning new heuristics and inhibiting such previously acquired inapplicable heuristics, then one may be able to explain many seeming puzzling aspects of development.

Thus, for example, Piaget's finding regarding the development of the concept of conservation of liquid can be explained using such a theory since in a middle class young child's world, the child's common experience would be that as more liquid is poured into her glass, the level rises. Such experience would naturally lead to the heuristic, the higher the level, the more liquid. Further, since the ability to inhibit responses is known to be under the control of the pre-frontal cortex which is the last portion of the brain to reach maturity (Luna, et al., 2001), it is not surprising that the development of true liquid conservation is protracted.

The brain and the sensory organs have evolved to function in the child's expected environment. Evolution is a stochastic theory in which traits that enhance the probability of survival and producing offspring are more likely to be passed on to the next generation. If not always the most important aid to survival, the ability to be able to predict what is most likely to happen next in one's environmental niche is surely important. As discussed previously, there is ample evidence that the brain utilizes a prediction feedback process in order to smoothly execute voluntary movements. As evolution is often described as conservative, that is, it co-opts existing genes and processes to create new functions, there is good reason to believe that the Prediction Process, Filtering Process and Inhibition Process working together with memory and

attention were cobbled together to enable us and other organisms to learn to cope with the vagaries of our environment by learning to recognize the stable or semi-stable aspects of our environment.

Endnotes

Chapter 1. Introduction and Background for the Theory of Prediction, Filtering and Inhibition Process in Learning and Memory.

1. Although possibly not relevant to the color identification question, there is a second corollary to the Prediction Process that is necessary to explain many psychological phenomena. If synaptic connections in the brain are modified through experience (a “Prediction Circuit”) to expect or predict a certain outcome (the “First Outcome”) if a certain set of conditions are met (and such conditions are usually met) but a different outcome can be expected (the “Second Outcome”) if all such conditions are not met (or if it turns out that the outcome predicted is totally erroneous as it is based upon a limited sample), then, in contradistinction to Piaget, there is no reason to believe that the original Prediction Circuit is itself modified. A more plausible explanation is that a second Prediction Circuit is established and the First Outcome that would otherwise be predicted by the first Prediction Circuit is inhibited if the brain is made aware of the changed conditions and the Second Outcome is predicted instead because the second Prediction Circuit is not inhibited (the “Inhibition Process”). See Luna, et.al. (2001) and Zelazo, Frye & Rapus (1996).
2. Connectionist simulations discussed later also show that while learning is dependent upon the initial state of the system, such connectionist systems will not only follow different trajectories as a result, but will also end up at different final internal states, yet the different final internal states all result in the same output (Salon, Second Year Project, 2000).
3. The ability to inhibit is tied to the maturation of the prefrontal lobes and may, together with the ability to communicate, account for the evolutionary advance in the cognitive power of the human organism (Luna, et.al., 2001).

Chapter 2. Learning Color Words

1. However, while the group data supported the hypothesis that pre-lexical concept organization supported language acquisition in the corresponding domain, the individual data was more varied and did not support the hypothesis as one-third of the non-color concept users learned in fewer trials than the average number of trials it took the color concept users to learn as a group.
2. It is unfortunate that many investigators have used the word constraint to describe these various word learning principles. The word constraint in

this context implies an inborn rule as well as a limitation. Yet in describing constraints the phrase “children expect” is often used and one can just as easily conceptualize the word learning constraints hypothesized by various investigators as subconscious rules or heuristics learned by children much as the rule for forming the past tense of English verbs by adding “ed” is learned by children along with the many exceptions to the rule. That is to say, these heuristics can be and are learned by children because they are useful in predicting the meaning of a word.

3. Smith, et.al. constructed a neural network to simulate the learning of color words. The network was trained successfully to answer questions equivalent to “What color is this?” “What texture is this?” and “What shape is this?” However, the hidden layer of nodes did not develop a common activation pattern for all red objects suggesting that the network had not abstracted a “meaning” for color. The architecture of this particular connectionist model was not a prediction architecture.

Chapter 5. Discussion

1. Of course, it should be noted that the children in the shape study were not tested to determine if they knew their colors. If they all did, then perhaps the difference could be attributed to this fact that they were generally more advanced verbally.

References

- Au, T.K., & Markman, E.M. (1987). Acquiring word meanings via linguistic contrast. Cognitive Development, 2, 217-236.
- Bearison, D.J. and Siegel, E.I (1968). Hierarchical Attributes for Categorization. Perceptual and Motor Skills, 27, 147-153.
- Berlin, B. & Kay, P. (1969). Basic color terms: their universality and evolution. Berkley: University of California Press.
- Best, C.T., McRoberts, G.W., LaFleur, R. & Silver-Isenstadt, J. (1995). Divergent developmental patterns for infants' perceptions of two non-native speech contrasts. Infant Behavior and Development, 18, 339-350.
- Bornstein, M.H. (1985). On the development of color naming in young children: Data and theory. Brain and Language, 26, 72-93.
- Chomsky, N. (1959). Aspects of a theory of syntax. Cambridge, Ma., MIT Press.
- Clark, E.V. (1987). The principle of contrasts: A constraint on language acquisition. In B. MacWhinney (Ed.), Mechanisms of Language Acquisition, (pp 1-33). Hillsdale, NJ: Erlbaum.
- Dambach, M., Rolfs, M., Gollner K., Kliegl, R. and Jacobs, A.M. (2009). Event-related potentials reveal rapid verification of predicted visual input. PLoS One 4(3): e5047. doi: 10.1371/Journal. Pone 0005047.
- Dess, N.K., Patterson, J. and Overmier B.J. (1986). Relative effectiveness of concurrent forward/backward versus simple forward and simple backward Pavlovian conditioning procedure. The American Journal of Psychology, 99, 1, 31-44.
- Diamond, A. (2000). Close interrelation of motor development and cognitive development and of the cerebellum and prefrontal cortex. Child Development, 71,44-56.
- Echols, C. (1991). Infant's attention to objects and consistency in linguistic and non-linguistic contexts. Paper presented at the biennial meeting of the Society for Research and Development, Seattle WA.
- Elman, J.L. (1990). Finding structure in time. Cognitive Science, 14, 179-211.
- Fellerman, D.J. and Van Essen, D.C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. Cerebral Cortex, Vol.1, pp. 1-47.

- Freeman, W.J. (1990). In John, E.R. (Ed.) Machinery of the Mind. Cambridge, Birkhaeuser, Boston.
- Gentner, D. & Ratterman, M.J. (1991). Language and the career of similarity. In S.A. Gelman & J.P. Byrnes (Eds.), Perspectives in language and thought: Interrelations in development, (pp. 225-277). Cambridge, England: Cambridge University Press.
- Gleick, J. (1987). Chaos: The making of a science. New York: Viking.
- Grossberg, S. (1999). The link between brain, learning, attention and consciousness. Consciousness and Cognition, 8,1, 1-44.
- Hawkins, J. and Blakeslee, S. (2004). On Intelligence. New York, New York: Henry Holt & Company.
- Hirsh-Pasek, K., Golinkoff, R.M. and Hollich, G. (2000). An emergentist coalition model for word learning: Mapping words to objects is a product of the interaction of multiple clues. In R.M. Golinkoff, et.al. (Ed.); Becoming a word learner: A debate on lexical acquisition (pp. 136-164). New York: Oxford University Press.
- Kemler-Nelson, D.G. (1995). Principle-based inferences in young children's categorization: Revisiting the impact of function on the naming of artifacts. Cognitive Development, 10, 347-380.
- Littschwanger, C.J., and Markman, E.M. (1994). Sixteen and twenty-four-months-old use of mutual exclusivity as a default assumption in second label learning. Development Psychology, 30, 955-968.
- Kotovskiy, L. & Gentner, D. (1997). Comparison and categorization in the development of relational similarity. Child Development, 67, 2797-2822.
- Krackow, E. and Gordon, P. (1998). Are lions and tigers substitutes or associates: Evidence against slot filler accounts of children's early categorization. Child Development. 69, 2, 347-354.
- Landau, B. & Gleitman, L.R. (1985). Language and experience. Cambridge, MA: Harvard University Press.
- Lowenstein (1999). The Touchstone of Life. New York, Oxford: Oxford University Press.
- Lucariello, J. Kyratzis, A. and Nelson, K. (1992). Taxonomic knowledge: What kind and where? Child Development, 63, 978-998.

- Luna, B., Thulboren, K.R., Munoz, D.P., Merriam E.P., Garver, K.E., Minshew, N.J., Keshavan, M.S., Genovese, C.R., Eddy, W.F. and Sweeney, J.A. (2001). Maturation of widely distributed brain function subserves cognitive development. Neuroimage 13,5, 786-793.
- Macario, J.F. (1991). Young children's use of color classification: Foods and canonically colored objects. Cognitive Development, 6, 17-46.
- Markman, E.M. & Wachtel, G.F. (1985). Children's use of mutual exclusivity to constrain the meanings of words. Cognitive Psychology, 20, 121-157.
- Maye, J., Werker, J.F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82, B101-B111.
- McDonough, L., and Mandler, J.M. (1998). Inductive generalization in 9- and 11-month-olds. Developmental Science, Vol. 1, No. 2, 227-232.
- Mountcastle, Vernon B. An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System (1978). In Gerdel M. Edelman and Vernon B. Mountcastle (Eds.). The Mindful Brain, Cambridge, Mass: MIT Press.
- Nelson, K. (1973). Structure and strategy in learning to talk. Monographs of the Society for Research in Child Development, 38, (1-2, Serial No. 149).
- Nelson, K. (1974). Concept, word and sentence: interrelations in acquisition and development. Psychological Review, 81, 267-285.
- Nelson, K. (1986). Event Knowledge. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Nelson, K. (2007). Young Minds in Social Worlds. Cambridge, MA: Harvard University.
- Newcombe, N.S. What is Neoconstructivism (2010). In Scott P. Johnson (ed.). Neoconstructivism: The New Science of Cognitive Development, Oxford, U.K.: Oxford University Press.

- Oxbury, J.M., Oxbury, S.M. & Humphrey, N.K. (1969). Varieties of color anomia. Brain, 92, 847-56.
- Paillard, J. Motion Determinants of a Unified World Perception. In Ascherslebern, G., Bachman, J. and Musseler, J. (Eds.). Cognitive contributions to the perception of spatial and temporal events. Chapter 5, 95-111. 1999 Elsevier Science B.V.
- Penrose, R. (1989). The emperor's new mind. New York, Oxford: Oxford University Press.
- Plunket, K., Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. Cognition, 38, 43-102.
- Praamstra, P., Steigman, D.F., Holstink, M.W.I.M., Brunia, C.H.M. and Cools, A.R., (1995). Movement-related potentials preceding voluntary movement are modulated by the mode of movement selection. Experimental Brain Research, 103, 429-439.
- Raichle, M.E. (2010). Brain's dark energy. Scientific American, March 2010, Volume 302 Number 3, New York, New York: Nature Publishing Group.
- Rice, N. (1980). Cognition to Language. Baltimore: University Park Press.
- Rumelhart, D.E., McClelland, John, and the PDP Research Group. Parallel Distributed Processing. Cambridge, Massachusetts/London, England: the MIT Press.
- Salon, J. (2000). A connectionist stimulation of the development of slot filler categories. Unpublished paper.
- Sandhoffer, C.M. & Smith, L.B. (1999). Learning color words involves learning a system of mappings. Developmental Psychology, 35, No. 3, 668-679.
- Saffran, J.R. (2010). What is statistical learning, and what statistical learning is not. In Scott P. Johnson (ed.). Neoconstructivism: The New Science of Cognitive Development. Oxford, U.K.: Oxford University Press
- Saffran, J.R. Aslin, R.N. and Newport, E.L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.
- Simon, H.A. & Newell, A. (1958). The Social Science Research Council, Items 12: 37-40.
- Skinner, B.F. (1969). Contingencies and Reinforcement. New York: Appleton-Century-Crofts.

- Slater, A. & Johnson, S.P. (1998). Visual sensory and perceptual abilities of the newborn: Beyond the blooming, buzzing confusion. In F. Simion & G. Butterworth (Eds.), The development of sensory, motor and cognitive capacities in early infancy: From perception to cognition, (pp. 121-141). Hove, U.K.: Erlbaum.
- Soja, N.N. (1994). Young children's concept of color and its relation to the acquisition of color words. Child Development, 65, 918-937.
- Soja, N.N., Carey, S. & Spelke, E.S. (1991). Ontological categories guide young children's induction of word meaning: Object terms and substance terms. Cognition, 38, 179-211.
- Sporns, O. & Edelman, G.M. (2008). Solving Bernstein's problem: A proposal for the development of coordinated movement by selection. Cognition, 64, 960-981.
- Tversky & Kahneman (1973 and 1974)
- Thelen, E.D., Corbetta, K. Kamm, J.P., Spencer, et.al. (1993). The transition to reaching: Mapping intention and intrinsic dynamics. Child Development. 64(4): 1058-1098
- Tomasello, M. & Brooks, P.J. (1999). Early syntactic development: A constructive grammar approach. In Martyn Barret (Ed.), The Development of Language, (pp. 161-190), Hove U.K.: Psychology Press Ltd.
- Tomasello, M. and Hubert, K. (2003). Understanding attention: 12 – and 18-month-olds know what is new for other persons. Developmental Psychology, 39(5), 906-912.
- Tversky, A. & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 5, 207-32.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-31.
- Werker, J.F. & Tees, R.C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development. 7, 49-63
- Wertheimer, M. (1938). Numbers and numerical concepts in primitive peoples. In W.P. Ellis (Ed.). A sourcebook of Gestalt psychology. New York: Harcourt, Brace.
- Waxman, S.R. and Markow, D.B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. Cognitive Psychology. 29, 257-302.

- Wolf, P. (1999). Space Perception and Intended Action. In Ascherslebern, G., Bachman, J. and Musseler, J. (Eds.). Cognitive contributions to the perception of spatial and temporal events. Chapter 3, 43-63. 1999 Elsevier Science B.V.
- Woodward, A.L. (1992). The role of the whole object assumption in early word learning. Doctoral dissertation, Stanford University, Stanford, CA
- Woodward, A.L., & Markman, E.M. (1998). Early word learning. In (Damon, W., Kuhn, D. and Siegler, R.S. (Eds.). Handbook of Child Psychology, 2, 371-420. New York: John Wiley & Sons, Inc.
- Zelazo, P.D., Frye, D. & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. Cognitive Development, 11, 37-63.