Incremental Repeated Acquisition in Children: Examining How Chain Type Might Influence the Association with Measures of Executive Function.

by

Kristen Spencer Walstrom

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the requirements for the Degree of
Doctor of Philosophy.

Auburn, Alabama
December 12, 2015

Keywords: Incremental Repeated Acquisition, Full Scale IQ, Dimensional Change Card Sort Task, Forward Chaining, Backward Chaining

Approved by
Chris Newland, Chair, Alumni Professor of Psychology
Jennifer Gillis, Co-chair, Associate Professor of Psychology
Martha Escobar, Associate Professor of Psychology
Sacha Pence, Assistant Professor of Psychology

Abstract

The Incremental Repeated Acquisition (IRA) procedure requires a participant to learn to produce a new sequence of responses each session. The current work makes two novel contributions to the small body of literature describing the use of an IRA procedure in humans. First two methods of chain development; Forward and backward chaining are compared. Second, this is the first time an attempt has been made to obtain more than a single IRA measurement with children. The study also provides an expansion of previous work regarding the association between IQ, age and performance on a Backward Chaining IRA task by including younger participants (2.5 to 7 years) and a measure of executive function, the Dimensional Change Card Sort task.

Previous research has indicated the performance on the Backward Chaining IRA task improves with age and increases in IQ. This study found that performance on both Forward and Backward Chaining IRA improved with age but not with IQ. The potential importance of various procedural differences between this and previous work is discussed. A within-subjects comparison of the Forward and Backward Chaining IRA indicated participants performed better on the Forward Chaining IRA Task.

Acknowledgements

“Thank you” seems a small word to acknowledge the effort and time of so many people who contributed to this dissertation and to my education. Still, it is with heartfelt gratitude that I acknowledge the following people for their generosity to me, and this project.

I would first like to thank my professors Chris Newland and Jen Gillis-Matson. Dr. Newland’s attention to my education has made me a better behavior analyst, and the time he has spent with me, not just on this project, has been invaluable. Dr. Gillis provided both practical support and encouragement and allowed me to retain my roots in applied work. In addition to all this I want to thank them both for their patience during my graduate school career. Thank you also to my committee, for their consideration and time in reading and commenting on this project.

Another important thank you goes to Andrew Kunzel, the Auburn alumni who wrote the computer program which formed the basis of this study. Throughout the duration of this project several other graduate students provided support as research assistants. I thank them heartily. I would also like to thank every participant, and their parents for allowing these remarkable children to work with me. I must also thank the various research locations who let an outsider in. The directors of these sights allowed me to take up time, space and solicit their students; all for no benefit to themselves. I am grateful.

I wish also to acknowledge my family; my parents Mark and Debbie, my sister Linzy and my husband Sean. It is impossible to enumerate all of the many ways my parents have supported me and my education over the years, they have been more than encouraging. My sister Linzy was a source of inspiration and novel ideas to solve the problems I encountered. My family was willing, at every turn, to roll up their sleeves and help in whatever way they could. From inviting participants to reading my work there was no task they turned from when I asked for help. Finally to my husband Sean, though I'm mentioning you last you are first in my heart. Thank you for walking with me on this journey, thank you for your pride in me and for never letting me give up.

And to God, to him be all the glory. Matthew 17:20

Table of Contents

List of Tables

List of Figures

Chapter 1 Literature Review

Examining how chain type might influence the association with measures of executive function.

Learning has received a great deal of attention in psychological literature and a variety of definitions exist. Here the definition of learning by Newland and Reile (1999) or Sidman (1960) was the primary basis for further discussion. Learning was defined "as the change in behavior from one steady-state to another in response to a change in the environment" (Newland & Reile, 1999, p. 321). This definition allows the behavioral scientist to treat learning as a phenomenon that can be reproduced and studied. Many laboratory tasks aimed at assessing learning only capture initial learning once, when the participant encounters the problem the first time. Repeated exposure to the task may test memory or proficiency but not the special learning that occurs during the first time an individual encounters the task. Even when behavioral transitions are arranged to occur repeatedly the apparent improvement in learning may in fact be the extinction of extraneous behaviors (Newland & Reile, 1999). This possibility emphasizes the need for a steady-state behavior in order to accurately examine learning processes.

One solution to the problem of learning that meets these criteria is the Incremental Repeated Acquisition (IRA) task. The task is uniquely suited to address learning as steady-state behavior because it allows the investigator to repeatedly examine the learning process as it occurs in the same individual. The IRA procedure is named based on a description of the task; the task provides an opportunity to *repeatedly* measure the *acquisition* of a new behavioral chain during the Learning condition (when a new sequence is introduced in each session). The behavioral chain is established *incrementally* over a session using either forward chaining methods or –more traditionally- backward chaining methods (Bailey, Johnson & Newland, 2010). IRA was developed by experimental psychologists in the animal laboratory to test learning using operant methods (Weinberger & Killam, 1978). The IRA procedure requires an animal, or human participant, to learn to produce a sequence of responses, a behavioral chain, during a session.

Although there are many procedural variations possible with the IRA task a typical arrangement used with humans to acquire a chain, designated A-B-C, is described as an illustration. Either forward or backward chaining methods can be used to establish a response chain. In forward chaining each new response is added to the end of the sequence. The individual is first required to emit a single response “A” followed by a reinforcer. This is the first response in the final behavioral chain A-B-C. When criteria are met, for example three consecutive correct responses, a second response “B” is now added to the sequence. The individual must now respond “A-B” before a reinforcer is available. After criteria are again met a third response “C” is added to the chain so that the individual must now respond “A-B-C” in order to access the reinforcer. In contrast, backward chaining begins with the final response in the sequence A-B-C and adds responses to the beginning of the chain. In this method “C” is followed by reinforcement. After criteria are met the chain increments and “B-C” is followed by reinforcement. Finally “A-B-C” is followed by reinforcement.

The current investigation examines the correlation between children’s performance on an IRA task, including a forward chaining variation, and IQ scores as measured by the Wechsler Preschool Primary Intelligence Scale. Two previous studies have identified a significant positive correlation (Paule, Chelonis, Buffalo, Blake & Casey, 1999 & Baldwin et al., 2012) with the backward chaining variation of the task. In addition another measure of executive function, and perseveration, will be used in an effort to identify other potential correlates of IRA performance. This review will first briefly examine the IRA task as conceptualized in the animal laboratory and developed from the repeated acquisition task. This will be followed by an examination of how the IRA procedure has been used in human populations and how the various aspects of chaining affect the procedure. Then the possible link between IRA and IQ first described by Paule and colleagues (1989; 1999) will be examined.

Early Development in the Animal Lab

Evolution of Repeated Acquisition

Incremental Repeated Acquisition (IRA) is a modification of the Repeated Acquisition task developed by Boren (1963) and Boren and Devine (1968) (Weinberger & Killiam, 1978). The IRA variation included elements from two other acquisition tasks (Pieper, 1976; Thompson, 1970) and was first described by Weinberger and Killam (1978). Both Repeated Acquisition and Incremental Repeated Acquisition have been widely used in the characterization of drugs and environmental contaminants. The tasks allow for the use of steady-state research methods to examine behavioral change in an individual animal over time. The evolution of the procedure is depicted in Table A1 and procedural variations used with animals are depicted in Table A2 (These tables are located in Appendix A). Detailed descriptions of several critical studies are available in Appendix C Part 1 page 102.

The principal modification of the repeated acquisition task, making it the IRA task of today, occurred in the late 70's with studies that incremented the chain length within the course of a single session (Piper, 1976; Weinberger and Killam, 1978). This differed from Boren's method (Boren & Devine, 1968) of developing a chain across sessions using shaping and establishing a high Fixed Ratio (FR) reinforcement schedule. This meant every session the animal started with a short chain that increased with accurate responding (21 consecutive errorless chains were required to advance the chain). The original IRA procedure (Weinberger & Killam, 1978) varies significantly from the critical procedural elements that have become associated with the IRA procedure. For instance, Weinberger and Killam did not use a performance or learning component, described later. In addition although Weinberger and Killam (1978) used backward chaining, they started animals at a two chain link and incremented the chain across sessions within a day (up to seven 15 minutes sessions in a day). By using an incrementing chain without explicit shaping procedures Weinberger and Killam (1978) combined the two most important procedural changes from Thompson (1970) and Pieper (1976).

The addition of an incrementing chain, essentially a non-explicit shaping procedure, sets the IRA procedure apart from repeated acquisition and allows for greater variety of procedural techniques. The behavioral chain used in repeated acquisition and IRA tasks may be presented in one of three ways. Whole chain presentation is the basis of the early repeated acquisition procedure; whereas incremental repeated acquisition (IRA) procedures may present the behavioral chain using forward chaining or backward chaining. Early IRA work makes use of backward chaining. A comparison of the two chaining options has only recently become available in work produced by the Auburn University's Behavior Pharmacology and Toxicology Lab.

Bailey, Johnson and Newland (2010) manipulated both the structure of the chain and the type of chain used to build a sequence while challenging the behavior of rats with d-amphetamine. They found no difference in performance on the two chain types (forward or backward chaining) across baseline or drug administrations. During baseline, no effect of chain structure was evident; however during drug administration, differences were obtained. These results indicated that PQ scores were always high for the performance sequence, but among animals trained using forward chaining the non-repeating sequences had higher scores than the repeating sequences while for those trained using backward chaining the effect was in the opposite direction. Additional work from this lab group include an examination of IRA performance in mice, BALB/c and C57BL/6 (Johnson, Bailey, Johnson & Newland, 2010) and IRA performance during chronic oral administration of haloperidol with olfactory stimulus differentiating the performance sequence and learning sequences (unpublished, presented as Spencer & Newland, 2010a). Johnson and colleagues (2010) found strain-dependent differences in response to chaining type, forward or backward; however, these differences did not meet traditional levels of significance. A main effect across session and evident in all dependent variables indicated higher scores were obtained by backward chaining groups. In the other study Haloperidol impaired acquisition of new IRA sequences, but no significant difference was detected between forward and backward chaining manipulations during a learning condition (Spencer & Newland, 2009).

Procedural Issues

A few of the many procedural variables possible in an IRA task are discussed here. The complexity of this task means that there are many experimental design choices and some have received relatively little attention, even though they might be relevant to the overall results.

Use of both performance and learning components. Boren and Devine (1968) established an early effort to differentiate between performance of a learned task and learning of a new task by defining some errors as *learning errors* and other errors as *performance errors*. This effort was expanded by Thompson and Moershbaucher (1979); their chapter on using repeated acquisition techniques to assess drug effects explicitly defined the procedural component variations in a repeated acquisition paradigm. Pharmacologists established two components that are particularly useful in assessing drug effects. The animal is trained under baseline conditions to an established criterion on a single sequence; this *Performance Component* gives rise to the so called *Performance Sequence*. The Performance Component is followed by a *Learning Component or Repeated Acquisition*

Component where the animal experiences a new sequence each session. Occasionally during the Learning Component the Performance Sequence is reissued. This allows the investigator to compare the results of the performance sequences to learning phase sequences and thus distinguish between impairments of new learning as opposed to impairment of previously acquired skills.

Stimuli arrangements to enhance responding. Stimulus fading or instructional stimuli received comparatively more attention in early studies. In most arrangements this can be thought of as analogous to a prompting procedure. In Boren and Devine (1968) the procedure involved two paired sessions. In the first session a single light was turned on directly over the correct lever, in the second session all three lights over the correct group of levers were turned on. This arrangement was compared to one in which the three non-specific lights were arranged. The results of this comparison indicated variation amongst the animals suggesting that for some animals the light served as important stimuli while for others these lights were relatively extraneous and even fraudulent instructional lights did not disrupt performance. This result calls into question the importance of the various intervening stimuli; but only in well trained animals. This effect was further demonstrated when the Auburn Lab conducted a *tandem chain component* (they removed all auditory stimuli throughout the chain) with well-trained animals and found behavior was notably not disrupted (unpublished, presented as Spencer, Bailey & Newland 2008).

Thompson and Moerschbaecher (1979) also compared two stimuli arrangements, one with a fading element. In the non-fading procedure the color of the keys (blue, red or yellow) changed after three correct responses signaling progress through the chain. In the fading condition only the correct key was lit up in the correct color; subsequently the non-correct keys were illuminated through six steps to full power, thus they were equivalent to the non-fading condition. The results indicated fewer errors were made under the fading condition. This procedure might be thought of as analogous to an *errorless learning* protocol.

Schedules of reinforcement. Also in need of consideration is the schedule of reinforcement. The schedules of reinforcement for correct responses in early studies (Boren & Devine, 1968; Thompson & Moerschbaucher, 1979) were various intermittent schedules of reinforcement (FR 5 or VR 3 etc.) while late studies generally use continuous reinforcement, FR1, (Cohn, Cox & Cory-Slechta, 1993; Bailey, Johnson and Newland, 2010; Johnson et al., 2010; Spencer & Newland, 2010a; Wenger, Schmidt & Davvison, 2004). Some later studies do use intermittent reinforcement schedules; for example Poling, Cleary, Berens and Thompson (1990) used an FR 5.

According to Thompson and Moerschbaucher (1979) an FR 50 schedule produced more errors and longer pausing when compared to an FR 5 schedule. It is important however to specify along with the reinforcement schedule the unit of behavior to be reinforced. In Thompson and Moerschbaucher (1979) the unit of behavior is the sequence; thus, early in the session a single response on the correct lever is reinforced, while later in the session four correct responses are reinforced. This convention is seen in Bailey et al. (2010), Spencer and Newland (2010a), Johnson et al. (2010) but was not the case in Boren and Devine’s study (1968) where an FR5 appears to refer to each element of the behavioral chain such that a correct response must be emitted five times before the animal proceeds to the next link of the chain, which also must be emitted five times and so on until the animal obtains a reinforcer (For example a four response chain with responses A, B, C and D; A, A, A, A, A, B, B, B, B, B, C, C, C, C, C, D, D, D, D, D, Reinforcer). This is significantly different even from other early repeated acquisition procedures (Thompson, 1970; Pieper, 1976) where an FR schedule is in place for the entire sequence.

Treatment of errors. Because the treatment of correct and incorrect responses is so critical in the analysis of behavior these findings are given further consideration here. First, in relation to errors, the use of any time-out following an error ranging from 1 – 240 seconds in length was more effective than no time-out at all and all time-outs suppressed errors in a stable manner (Boren & Devine, 1968). What was not manipulated and has changed substantially in later versions of the IRA task is the decision to reset the sequence after an error. In general early versions of both repeated acquisition (Boren & Devine, 1968) and IRA (Weinberger & Killam, 1978) do not reset the sequence after an error. In contrast, later work (Bailey et al. 2010; Johnson et al., 2010; Spencer & Newland, 2010a) resets the sequence after an error requiring the participant to start the sequence over again and produce a sequence with no intervening errors in order to obtain a reinforcer. This difference might also be traced to the type of repeated acquisition procedure used. In repeated acquisition procedures with whole chain presentation a requirement to reset the chain after an error would alter the chain to essentially be forward chaining without preset advancement criteria.

Chain structure. Three studies have provided relevant information on the importance of chain structure and two of these included direct manipulation of the chain structure. Cohn, Cox and Cory-Slechta (1993) noted that learning chains which were more similar to the performance chain had better accuracy scores than those that didn't. A specific example is with a chain CLR (Center – Left – Right), repeated acquisition chains LRC and RCL had substantially higher accuracy scores than other repeated acquisition chains. Another study (Wright & Paule, 2007) specifically examined the difficulty of the response sequence in the IRA procedure. Using backward chaining up to a six-link sequence on three levers the study examined 16 different response sequences for difficulty. The results indicated that various chains do have different levels of difficulty and this difficulty appears related to the number of levers needed to complete the chain and the location of the levers within the chain. Chains using two levers were easier than those using three levers and those with adjacent responses were easier than those without adjacent responses.

Bailey and colleagues (2010) also specifically manipulated successive repetition of responses within the chain and found that for the forward chaining group there was no detectable difference between sequences that contained a repetitive link and those that did not. In contrast, a difference between the performance chain and non-repeating chains was detected in the backward chaining group, indicating that non-repeating chains were more difficult. In terms of chain selection, the structure is important.

Measurement issues. The selection of dependent variables has undergone modification like the technique itself (See Table A1). Early studies (Boren & Devine, 1968) used the number of errors as the primary measure of interest. While other studies (Pieper, 1976) instead used chain length attained. However, both of these measures have been criticized by Thompson and Moerschbaucher (1979) when they are used alone. In this context the problem with chain length, or even number of errors or percent error as a measure, is the lack of within-session information provided. Thompson and Moerschbaucher specifically request information on the acquisition of each response in the sequence in order to evaluate that acquisition has occurred. This view emphasizes that repeated acquisition procedures are designed for individuals and group data is inappropriate.

More recent work from Auburn University (Bailey et al. 2010; Johnson et al., 2010) has also struggled with the issue of appropriate measurement, precisely because grouped data is often necessary. Bailey, Johnson and Newland (2010) describe accuracy alone as an insensitive measure because it cannot differentiate between an animal that performs accurately on many short chains from one that performs accurately while reaching longer chains. When the number of reinforcers is allowed to vary at each chain length neither the use of reinforcers or percent task complete (Johnson et al. 2010) solve this problem. For example, one animal might attain many reinforcers on a short chain while another receives many reinforcers on a long chain. An additional problem with accuracy is encountered when forward and backward chaining styles are compared. Because backward chaining animals start with a novel response each time the chain increments, the individual is at a disadvantage in collecting accurate responses compared to an individual on a forward chain where a new chain starts with a previously mastered response (Bailey et al. 2010). Chain length is also insensitive, but here only to smaller differences between animals performing the same length of chain. These investigators (Bailey et al. 2010, Johnson et al. 2010) selected three measures with which they evaluated learning and performance: total responding (number of responses per session), accuracy (correct responses, even those that occurred before an incorrect response which reset the sequence, divided by Response Total); in addition this study introduced a novel dependent measure. The progress quotient or "PQ" score weights reinforcers obtained by the chain length and normalizes this measure by the total reinforcers obtained. The numerator can also be viewed as a count of all responses that formed correct chains. The formula for the PQ score is shown in the box for Formula 1 (below) where Rf_i = the number of reinforcers earned on a chain of length i and Rf_{tot} = total reinforcers earned in the session. In this study, a 4 link chain was the maximum possible; however the formula could be expanded to evaluate longer sequences.

Equation 1: Progress Quotient

$$\frac{(1*Rf_1 + 2*Rf_2 + 3*Rf_3 + 4*Rf_4 + \ldots + i*Rf_i)}{Rf_{tot}} \qquad (1)$$

According to these investigators the major drawback of the PQ is that it does not directly assess errors. Also, because it is specific to task manipulations it can be difficult to compare across studies. The PQ does allow investigators to compare actual performance to 'idealized performance' (Johnson et al., 2010). The PQ provides a more detailed picture of within session behavior without requiring multiple dependent variables which can be cumbersome as chain length requirements increase. The PQ also allows for comparison across subjects and allows group data in a way that complex multiple measures make difficult when several independent variables are included in a study.

The Adaptation of Repeated Acquisition Tasks for Humans

The trend in research from repeated acquisition to IRA has not been equally strong with humans. A dozen or more published studies use repeated acquisition procedures in humans (Higgins, Woodward & Henningfield, 1989). Repeated acquisition procedures have actually been used more widely than IRA procedures in human populations to examine a variety of pharmacological substances (barbiturates, benzodiazepines, d-amphetamine, alcohol and cocaine) as well as non-pharmacological studies (Higgins, Woodward & Henningfield, 1989). In contrast the IRA procedure has been used very infrequently in human populations.

The history of the IRA task in humans is summarized in Table A3 and various procedural variables are included as well. The first time investigators attempted to bring the IRA procedure out of the animal lab and into the domain of human operant research was in the late 80's. Investigators at the National Center for Toxicology Research (NCTR) developed the Operant Test Battery (OTB) in an effort to create a method of screening for toxicity using behavioral tasks (Paule, Schulze & Slikker, 1988). The OTB included five operant tasks selected to represent important human functions. The IRA was one of these tasks and was included to assess learning. Between 1988 and 1999 the NCTR investigators made an early effort to explore the OTB in a human population, publishing three studies and a book chapter in the process. The NCTR investigators justified their decision to expand the OTB to humans by identifying the importance of improving risk assessment in humans. Although the NCTR investigators were interested in the utility of the entire OTB, their publications also represent the bulk of existing human data using an IRA procedure.

The first use of the OTB with humans by Paule and colleagues (1988b) sampled a group of twenty children between the ages of 3 and 11 years. In all OTB validation studies the same panel originally designed for Rhesus monkeys was used. This was a large operant instrument panel with three ‘press plates' (flat response instruments), four retractable levers and various lights and speakers. The panel was installed in a research room at a local hospital and was modified to dispense nickels to children instead of the food pellets used in the animal lab. One important feature of the NCTR administration of the IRA task was the presence of verbal instructions and a demonstration via a video recording (please see Appendix B for a copy of the instructions). The instructions were 354 words long and had a reading grade level –as text assessed by Microsoft Office Word- of 6.4; to be clear participants didn't read the instructions, the Microsoft Office Word assessment is provided to compare direction across different studies. This group of children received an average of 16.2 reinforcers +/- 0.8. In this task it was observed that most children could not complete the whole task and the ability to complete the task appeared to be age-related and also related to clinical diagnosis. The results showed that in children over six years of age, only those diagnosed with ADD or a “Learning Disability” (this study didn't use current conventions) were unable to complete the IRA task. The investigators didn't report accuracy data. A second study from this group is included in Appendix C Part 2.

The third study published by Paule and colleagues (1999) focused on examining correlations between traditional measures of cognitive ability, namely IQ testing and the OTB component tasks. In this study 115 low birth-weight preterm 6 year old children enrolled in a collaborative longitudinal study on educational practices were tested using the OTB. The outcome data were displayed based on six IQ groupings, accuracy ranged from an average of 33% for 14 low IQ participants (Full Scale IQ < 70) to 63% for 12 high IQ participants (Full Scale IQ > 110). The investigators used Wechsler Preschool Primary Scale of Intelligence (WPPSI) IQ scores collected a year prior to the OTB component testing when all children were 5 years of age. Correlations were calculated from 92 of these participants for Full Scale IQ, Verbal IQ and Performance IQ and the three IRA outcome variables: accuracy, percent task complete (defined as the number of reinforcers obtained out of the total possible reinforcers, 18 for this task) and response rate per second. There is no discussion of why only 92 of the 115 total participants were used in calculating the correlations. The correlation between IRA accuracy and Full Scale IQ, $r=0.53$ ($n=92$), was significant and quite high compared to other correlations with OTB tasks. Only the correlation between accuracy on the Conditioned Position Responding (CPR) task was higher, $r=0.58$ for Full Scale IQ ($n=107$). The next highest correlation was for the Delay Match to Sample task, Full Scale IQ and accuracy $r=0.44$ ($n= 99$). Additionally, both VIQ and PIQ showed significant correlations with accuracy on both the CPR task (VIQ $r=0.516$ $p<0.01$ $n=107$; PIQ $r=0.569$ $p<0.01$ $n=107$) and the IRA task (VIQ $r=0.461$, $p<0.01$, $n=92$; PIQ $r=0.516$, $p<0.01$, $n=92$). These results were used to suggest that the operant tasks designed in the animal lab can assess important human brain function and provide information that may not be obtained from traditional assessments of intelligence. Further, the correlation between IQ and the IRA task may suggest a more specific relationship compared to other OTB tasks.

A more recent study by Zayac and Johnston (2008) used the IRA task to capture and contrive establishing operations during learning. Of note, although Zayac and Johnston were also at Auburn, their work was completely independent from the series of projects completed in the Newland and Gillis labs. Zayac and Johnston (2008) used more current technology as a platform for the IRA procedure, a desk-top computer with an add-on touch screen (visual basic). The procedure was designed around testing performance given various deprivation phases from the reinforcer, 40 second access to a participant selected computer game, ranging from fifteen minutes to three days of deprivation from video games. The procedure was designed based on Paule et al. (1988), but with several adaptations. Although the paper cites Paule et al's (1988) study, it does not explicitly describe the incrementing method. It is presumed that backward chaining was used but is not clear from the description given. Three participants were selected from individuals who attended an adult day care program; all were males previously diagnosed with mild to moderate Intellectual Disability (ID) between the ages of 37 and 43. The number of errors and percent of errors were presented as results, errors per session ranged from 55 to 341 and varied reliably based on the period of deprivation. The percent errors for all participants at all chain lengths, ranged from 28% to 40%.

A series of three studies was completed by Spencer Walstrom, Gillis and Newland (2011, unpublished data). These studies use a desk-top computer, add-on touch screen and REALbasic to create and conduct an IRA program. The project was completed with undergraduate college students at Auburn University. This study expanded the history of IRA with humans by using both forward and backward chaining methods to develop the behavioral chain throughout a session. In addition, the first study specifically aimed at determining if a performance like component is necessary or if a single measurement (or immediate exposure to a learning phase) adequately measures ability on the IRA task. The second study attempted a delayed follow-up, a second set of measurements between participants. A final study compared response modality, responding on a touch screen versus a mouse, and the pacing of setting and consequent stimuli. The program used auditory tones to establish the chain schedule between responses. Like previous studies (Paule et al. 1999), three correct responses were necessary to add a step to the behavioral chain (in this case consecutive correct responses were required). The sessions lasted for various lengths of time (one 40 minute session in study one, two 20 minute sessions in study three), with no maximum on the number of reinforcers a participant could receive (A screen reading, “Good Job, 100 points!). The program showed four square blue response options in a symmetrical square cluster. The response options were only differentiated from each other based on position. The program allows a maximum of a nine-link chain. Each individual was issued the following simple instructions, “You can earn points by pushing some of the different buttons you see. Try to earn as many points as you can. Please only respond by touching the screen, don’t use the mouse. Also, please do not use your cell phone during the study,” (44 words, 3.4 Flesch-Kincaid grade level as assessed by Microsoft word). The instructions were manipulated slightly based on the variation of the study.

Across these studies a ceiling effect was noted, participants almost all completed the entire chain with an extremely high degree of accuracy. The general findings suggest that a performance like component makes little difference in the chain length or accuracy attained by college students. In general individuals in the forward chaining group outperformed those in the backward chaining group except when participants were invited back to the lab after a delay; then participants in the backward chaining group slightly outperformed those in the forward chaining group; however, the utility of this finding is limited as not all participants chose to return to the lab for the follow up. No differences were found based on response modality. Stimulus pacing did affect the rate of responding but not measures of performance such as accuracy or length of the behavioral chain achieved.

The latest study looking at IRA performance in humans was published in 2012 by Baldwin, Chelonis, Prunty & Paule. This study is the latest work from the NCTR to expand the understanding of their OTB in a human population. In this study, 837 children completed the entire OTB, but the results presented here focus on the IRA task. This large scale study was conducted in an effort to obtain a large normative sample from typically developing children. The study also aimed to evaluate if sex differences seen in primates were detectable in humans and if the IRA continued to be sensitive to IQ differences in children of various ages, 5 to 13 years old. This study appears to use a procedure very similar to that developed by Paule et al. 1988; however, quite a bit more detail about the procedural details of the IRA task is presented. It is not clear if this detail is presented for clarity purposes, or because it represents departures from early NCTR studies. These details, like how errorless chains were defined (errorless excluding the most recently added response) and the error correction procedure (errors did not re-set the chain, and a nickel was always obtained at the end of the sequence if corrections were made) may become particularly important when the results of this study are compared to others. In this study the Kaufman Brief Intelligence Test (Kaufman & Kaufman, 1990) was used instead of the WPPSI.

The results of this study confirmed that participants' IQ was significantly predictive of IRA performance, particularly in younger children, but these differences in IRA performance were attenuated as age increased. Age also appears to be a particularly strong predictor of performance on the IRA task. The investigators noted that chain length was a limiting factor among older children, where a ceiling effect was evident and the consequent restriction of range could have prevented the detection of a relation. This was not as much of a problem with younger children who were often unable to advance to the longer chains.

Measurement issues. IRA results from humans have not been presented in a consistent fashion. Paule and the NTCR (1988b, 1990b, 1999) utilized accuracy, percent task complete and the response rate per second to summarize IRA task behavior. Zayac and Johnston (2008) used percent errors and errors per session as the primary dependent variable but did also calculate accuracy and responses per minute. In contrast Spencer Walstrom and colleagues have used Response Total, accuracy, PQ (progress quotient as defined by Bailey, Johnson and Newland, 2010) and chain length to examine IRA task behavior. Baldwin et al (2012) used the traditional NCTR test measures but added an analysis of responding based on *effective responses* (correct responses distinguished from ineffective correct responses occurring during various timeouts) within the session; "Search" responses (responses occurring prior to and including the first correct response each time the chain incremented) and "Memory" responses (responses occurring after the first correct response at each new chain length). In addition a *response chain accuracy* variable was calculated, as opposed to a simple accuracy value. Several of the NCTR test studies also present data at various chain lengths (two link chain, three link chain and so on) in accordance with early repeated acquisition and IRA work in various animal labs.

Features of Chaining as a Measure of Learning

The dominant behavioral perception of the response chain is of a sequence of behaviors that can be defined as a succession of different operant responses each reinforced by producing an opportunity to engage in the next response until the behavior chain is complete, terminated by a reinforcer (Catania, 2007). This creates a pattern where response *members* are *linked* by stimuli which serve as both a discriminative stimulus (Sd) and a conditioned reinforcer (Millenson, 1967). In most current texts (Catania, 2007; Martin & Pear, 2007) chaining is depicted as a linear process; however, in older textbooks chaining is depicted as a cycle. This depiction is apt in light of the way IRA works within the laboratory and may be important given that applied literature contributes the most information on the comparison of chaining strategies.

Although this definition of chaining focuses on the presence of conditioned reinforcers throughout the chain, Lashley (1951) identifies chains that are not maintained by association. These examples include the gaits of a horse as well as the finger movements of a musician. It seems that this dichotomy might be partially addressed by the examination of function within the chain. For instance each step in the gait of a horse contributes to the same function, whereas in the example of a response chain given by Sidman (1960, p. 101-103) and used by Catania (2007) the chain learned by a lever pressing rat can be broken down such that the lever pressing is extinguished but other aspects of the chain remain intact (approaching the food bin when non contingent food pellets are released etc.). The general conclusion is that a behavior chain can be discriminated on the ability to break it down into individual operants; or by our conception of the chain as a temporally extended unit of behavior.

Comparing the Techniques of Forward Chaining and Backward Chaining

As previously noted there are two methods of developing a behavioral chain in an incremental repeated acquisition procedure; either forward or backward chaining. When forward and backward chaining are considered simultaneously an initial theme is the question: Does one of these opposing techniques teach the behavior chain “better” or in effect establish a more or less robust behavior chain? This first question, of superiority, is followed by consideration of the different behavioral contingencies at work within each chaining method. References to lab lore regarding a preference for backward chaining among behaviorists (Bailey et al. 2010), given its perceived superiority, likely arise from a variety of sources published in the 60’s and 70’s promoting the method (Pisacreta, 1982 & Weiss, 1978). The theoretical underpinnings of this preference will be briefly explored after the initial question of technical superiority is examined.

Direct comparisons offering an empirical evaluation of the two methods do not conclusively favor one method over the other. These studies evaluate backward and forward chaining methods in several different participant populations in order to teach many different tasks. In several cases no differences can be detected. A brief review of studies offering a direct comparison of the methods is presented in Table A4, details regarding the studies selected are in Appendix C Part 3.

These studies were analyzed first based on the outcome measures used to determine supposed superiority of a chaining method. Of these 13 studies 11 contained an analysis of chaining based on some measure of performance during acquisition, for instance number of errors or accuracy. Six of these studies found no difference between chaining types, four studies suggested forward chaining was superior and two suggested backward chaining was superior. Four studies contained an analysis of chaining based on the length of time required during training to establish the chain, such as duration of the session, number of trials to criterion etc. Of these studies three showed no difference between forward and backward chaining and one suggested that forward chaining was superior. Finally four studies examined outcomes based on retention of the skill taught with a retest window ranging from immediately, 20 hours post training to one week post training. Of these studies three indicated no different and one suggested forward chaining was superior. In general regardless of the outcome measure it was most common to find no difference between forward and backward chaining; although, individual cases of forward and backward chaining appearing superior do exist. Given the lack of a clearly superior technique the initial question regarding superiority transitions into a consideration of when the selection of one technique over the other is warranted, in this case it becomes valuable to consider the participant and the type of skill targeted.

The apparent preference for backward chaining in behavior analytic literature isn't supported empirically, but where did this preference arise from? Several older texts don't even describe forward chaining as a method useful in developing a chain (Millenson, 1967; Rachlin, 1935, Ferster & Perrot, 1968). According to Millenson (1967) each stimulus response pair must be established as a discriminative operant first and then the Sd may serve as a conditioned reinforcer for the next response to be added to the chain. This effectively builds an association between each response in the chain and the reinforcer. This association would make each successive behavior in a chain more reinforcing until the reinforcer is attained at the end of the chain. The dual role of the linking stimuli throughout the chain intersects with a discussion below regarding what conditions allow a discriminative stimulus to also function as a conditioned reinforcer.

In contrast to the associations built in backward chaining; forward chaining explicitly and directly reinforces each new link of the chain as it is added (Weiss, 1978). The theoretical problem with forward chaining is that with the addition of each new “final” link the previously established behaviors are no longer directly and immediately reinforced, risking extinction of the established behavior (Catania, 2007). What is seldom considered in the context of forward chaining is all the possible side effects of extinction; Neuringer (2000) has identified extinction, or adversity by withholding reinforcement as a potential source of variability in behavior. Given a brief period of extinction which does not overwhelm the previous history of reinforcement, or experience with intermittent reinforcement, we can see how variations in behavior could produce the newly selected behavior and generate a lengthening chain.

The role of linking stimuli within the chain is another potentially important theoretical difference between forward and backward chaining as part of the IRA task. In a developing behavioral chain the dual role of these stimuli – discriminative stimulus and conditioned reinforcer - might even be expanded to a third role as an informational cue that simultaneously signals that another response is necessary and/or the proximity of reinforcement in terms of the quantity of responses required to obtain a reinforcer. It is easy to conceive that these stimuli might function differently in a forward and backward incrementing chain. In a forward chaining IRA task during the first trial in which a new link is added to the end of the chain the participant is most likely unaware of the alteration in the reinforcement schedule until they are most of the way through the new chain. In contrast, during the first trial of a lengthening backward chain the participant is likely immediately aware of an altered contingency.

The chain of associations described in backward chaining also applies to the discriminative stimuli present in a *chain schedule* as well (one where a discriminative stimulus intervenes between each response). Gollub (1958 as cited in Fantino, 2008) specifically compared a *tandem schedule* to a chain schedule, and found that the chain schedule did not support high rates of behavior, in fact over time behavior under the chain schedule was markedly reduced. Fantino (2008) describes this effect, noting that despite the supposed conditioned reinforcement at each link of the chain schedule only the final stimuli is ever directly paired with the primary reinforcer. In contrast under the tandem schedule, where a single stimuli is present throughout the entire chain, the stimuli present is always directly paired with the primary reinforcer. The tandem schedule more closely matches forward chaining in the IRA task. In the forward chain although the discriminative stimuli are different at each link of the chain (a chain schedule) for at least some short period each stimulus is also directly linked with a reinforcer; thus; forward chaining is able to generate considerable behavior. In the context of backward chaining, it seems that Gollub’s difficulty with the chain schedule ought to be a problem; in practice however behavior is maintained on a backward chaining IRA schedule. Through a rigorous review we see that conditioned reinforcers are only those correlated with a reduction in time to primary reinforcement (Fantino, 2008). Interestingly; under a choice paradigm set up to mimic Gollub’s arrangements pigeons preferred, as interpreted by selection, the tandem schedule. Other work also showed that under choice conditions pigeons preferred conditions without conditioned reinforcers; despite the fact that conditions with and without the conditioned reinforcers supported behavior (Schuster, 1969; Squires, 1972 as cited in Fantino, 2008). This preference may be explained by the demonstrated disassociation (Gollub, 1958; Fantino 1965) of the discriminative stimulus-conditioned reinforcer relationship (Keller & Schoenfeld 1950; Skinner, 1938). It is probably incorrect to view the discriminative stimuli throughout the IRA chain as automatically serving as conditioned reinforcers.

If we conceive the various discriminative stimuli throughout the IRA chain (a chain schedule, not a tandem schedule) as informational, in any way, we must consider *observing behavior*. Observing is the behavior exhibited by an animal who may respond to obtain access to a stimuli which identifies if the animal is in a reinforced contingency, or not, but this response has no effect on that contingency (Wyckoff, 1952). The behavior is well established in various species (Fantino, 2008) as is the specific condition in which it occurs; observing only occurs in relation to positive reinforcement (Dinsmore, 1983; Fantino, 1977). In these cases high rates of observing behavior can be observed even during extinction conditions, when behaviors effective in producing a reinforcer are suppressed. It is not clear if the type of information indicated by the stimuli within the IRA chain, information about the proximity of reinforcement, is functionally similar to information about the availability of reinforcement. In the context of backward chaining the idea of intervening IRA stimuli serving as informational cues throughout the chain intersects with the concept of ratio strain.

It is known that informational cues which indicate an extinction condition suppress behaviors, but what if the cue only indicates a small increase in the effort required to obtain the response? In the IRA task the reinforcer ratio is effectively a fixed ratio one schedule which graduates with the length of the chain, with a maximum fixed ratio schedule varying from four to nine (as conducted in Bailey, Johnson & Newland, 2010; Johnson et al. 2010; Spencer & Newland 2011, unpublished). The increase in effort involved in these extending chains may be viewed differently from a schedule in the pigeon lab which might support 100’s of responses per minute. If the information provided throughout the sequences is different for forward and backward chaining we might expect to see some interaction with the participants' motivation.

Another primary difference between forward- and backward-chaining is the repetition of previously-learned responses prior to each new response in a forward-chaining procedure. This difference leads to a discussion of how an individual's perseverative tendencies might help or hinder the individual during the acquisition process. Perseveration, the tendency to repeat behavior without regard to the consequences, might be viewed as particularly detrimental to the development of a backward chain where each new response is added to the beginning of the chain. In contrast forward chaining might mask preservative tendencies by allowing previously learned responses to occur first and the new response to occur at the end of the chain. Because this possibility has not been directly assessed before we will examine methods of measuring perseveration, in particular a task designed for children.

A Deeper Look at the IRA Executive Function Link

IQ information. IQ testing has a long history and will only briefly be described here. Paule and the early NCTR investigators made use of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI, 1974 edition) so our discussion will focus primarily on the WPPSI–III, the current version. The WPPSI-III consists of a series of 14 subtests for children age 4 to 7 years 3 months, only 7 of which are core tests. For children between the ages of 2 years 3 months and 3 years 11 months there are 5 subtests. The longer version takes approximately 45 minutes to complete. In general the use of prompts and queries is not restricted (Pearson, 2012).

Neyens and Aldenkamp (1996) examined the stability of the WPPSI-III (and several other measures) in children between 4 and 13 years of age. They report that typically a two-year window between testing is used to ensure that changes are not merely the result of practice, but this is too long a window to be useful in assessing changes in children with neurodegenerative disorders. Stability coefficients are reported from Razavieh and Shahim (1990) for FSIQ, 0.83, VIQ, 0.67, PIQ, 0.87 which had a test-retest interval of less than 40 days. Neyens and Aldenkamp used a six month test-retest window between three assessments and found excellent reliability for the PIQ and the FSIQ although all three IQ values changed across the three assessments. VIQ changed the most between the first and third assessment, 1 year interval, indicating gains in vocabulary and comprehension. The specific nature of improvement in the VIQ subtests gives way to questions about the language based nature of the WPPSI-III. Stark, Tallal, Kallman and Mellits (1983) found that when specifically language delayed children's performance on the WPPSI-III were compared with a matched sample of typically developing children outcomes on the Performance IQ subtests were not significantly different between-groups.

The stability of the IQ assessment results is particularly important given the difference between the early NCTR work (Paule et al. 1999) and the proposed study. In that NCTR study the IRA is correlated with Full Scale IQ scores that are over a year old. In the proposed study the IQ scores will be taken within two weeks of IRA assessment. Additionally all children participating in the NCTR data were given the IQ test at the same age (5) while the children in this study will be given the appropriate subtests at various ages when they experience the IRA assessment. Baldwin et al. (2012) used the Kaufman Brief Intelligence Test (K-BIT, Kaufman & Kaufman, 1990) to obtain measures of IQ. In this case the Verbal and Matrices subtests of the K-BIT were presented immediately after OTB testing, approximately 20 minutes of testing.

The NCTR paper described what they called a strong correlation between the IRA procedure accuracy scores and individual IQ scores, p=0.53 (Wechsler Preschool Primary Scale of Intelligence, WPPSI-III; Paule et al. 1999). This significant correlation was unique to IRA accuracy, backward chaining, and CPR among the OTB component tests. Although their validation was encouraging and provided a meaningful corollary to animal models; the OTB and its component tests have remained virtually unused in human testing. The OTB component tasks were analyzed in relation to each other based on animal data (Rhesus monkeys). This analysis showed that the correlation between tasks was generally low (less than 0.5, ranging from 0.02 to 0.39) and non-significant, indicating that the tasks measured relatively independent functions. Between the IRA task and the CPR, the tasks with the highest correlation to IQ scores, a non-significant correlation of 0.213 was detected (Paule, 1990c). It is important to note that the CPR (Conditioned Position Responding, a task requiring the acquisition of a color-response choice discrimination) task actually has a higher correlation with IQ scores than IRA; however, since this measure is focused primarily on visual discrimination it has not been targeted for additional examination with regard to a link with IQ scores.

Baldwin et al.'s (2012) normative sample also detected a relationship between IQ and IRA performance. In a two-way ANOVA with IQ and age significant main effects of IQ were detected for all four dependent variables, tested individually [Percent Task Complete (PTC) and IQ $F(2,784)=60.56$ $p<.01$; response chain accuracy and IQ $F(2, 784)=74.92$ $p<.01$; search response accuracy and IQ $F(2, 784)=8.70$ $p<.01$; memory response accuracy and IQ ($F(2,784)=56.08$ $p<.01$]. Participants were divided into three IQ groups (range of 70 to 131 as measured by the K-BIT), using .67 standard deviations from the mean (10 points); 71-90 below average, 91-110 average, 111-130 above-average. Across all age groups and dependent measures participants in the above average group outperformed those in the average and below average groups and participants in the average group outperformed those in the below average group. Search response accuracy was the only variable where this pattern is only moderately observed. Significant interactions between age and IQ were detected for PTC [$F(16,784)=2.27$ $p<.01$], memory response accuracy [$F(16,784)=1.85$ $p<.02$]

which indicated that among younger children IQ differences were particularly strong and this relationship decreased among older children. These investigators determined that *response chain accuracy* (defined as the number of times a response chain was completed correctly divided by the total number of times the subject completed a response chain) was more sensitive to IQ differences than percent task complete.

The relationship between IRA and IQ scores is only one example of the attempt to put the IRA into context as a lab task that depicts meaningful cognitive function. These attempts are designed to bridge the human- and animal-testing literature by using a test that is common to both. Frequently, but not always, these correlations are conducted with both full-scale IQ and subtests. The conception of repeated acquisition tasks as a measure of cognitive functions was also present in Shannon and Love's (2004) work. In this case repeated acquisition was specifically used as an executive function task. Executive function refers to a series of processes, including; planning, hypothesis generation, cognitive flexibility, decision making, judgment and feedback utilization. The Wisconsin Card Sort Test (WCST, also discussed in the next section on perseveration) is one common measure of executive function. These authors describe the repeated acquisition task as 'rule' learning across sessions and attempted to create an animal model of executive function comparable to that measured by the WCST. This was accomplished by placing animals on a within session repeated acquisition schedule, where a two response chain was required after 10 correct responses an un-signaled change occurred and a new two link chain was required.

Executive function. In a broad review Jurado and Rosselli (2007) described executive function as a frequently used concept that "….still awaits formal definition. (p. 123)" The review identified eleven different studies prior to 2007 which described over thirty components of executive function. Despite the expansive nature of the term executive function the reviewers chose to focus on four primary components of executive function: Attentional control, planning, set-shifting and verbal fluency (this paper will follow that practice and focus on the two functions that are relevant in the context of the IRA: Attentional control and set-shifting).

Attentional control was initially defined by Deerberry and Rothbart (1988) as the ability to voluntarily focus attention and the ability to shift attention based on the demands of the environment. This ability includes selective attention, sustained attention and response inhibition (Jurado & Rosselli, 2007). Attentional control is often characterized and measured by inhibition tasks like a Go/No Go task; with developmental improvements and a reduction in perseverative errors appearing between 9 and 12 years. Attentional control seems particularly relevant to one variation of the IRA task, the backward chain. With each added link of the chain there is a conceptual attention shift to a new stimulus, and inhibition of previously reinforced responses is necessary, at least temporarily. In this sense early repeated acquisition and backward chaining IRA tasks can accurately be described as requiring executive function.

Set shifting is often characterized as cognitive flexibility and is characterized by tasks that required rapid switching between response sets (Jurado & Rosselli, 2007). Set shifting is typically measured by a card sort task (such as the DCCS described below) with a response to verbal rules (i.e. red cards go here, blue cards go there…) but behaviorally it would be incorrect to rule out more discrete discriminative stimuli, like those existing between the responses of the IRA behavior chain as signals requiring a set shift. What is less clear is if the response chain that can be broken down into separate operants in fact makes up distinct response sets. Set shifting is discussed more below in the context of card sort tasks.

Card sort tasks. The classic test of perseveration developed from a cognitive orientation is the Wisconsin Card Sort Task (WCST), developed based on experiments in the Wisconsin Primate

Laboratory (Berg, 1948). This task conceptualizes perseveration as an inability to “set shift” with a major goal of quantifying the perseveration. Prior to the WCST, tasks were designed to examine perseveration, but few provided a quantifiable measure (Berg, 1948). The WCST is characterized as a test of cognitive flexibility, but there is disagreement among cognitive investigators with regard to what other cognitive functions WCST might measure (Geurts, Corbett & Solomon, 2009). Since the inception of the WCST several updated versions have been developed; the Modified Card Sort Task (MCST) and the Cambridge Neuropsychological Test Automated Batteries’ (CANTAB) intra-dimensional extra-dimensional (ID/ED) set shift task. The WCST has also been identified as a measure of executive function (Shannon & Love,2004).

The Dimensional Change Card Sort (DCCS) task is designed to measure executive function in children (Zelazo, 2006). This task is described in more detail in the Method section as it is a measure used in this study, but is also introduced here. The task consists of three phases (a total of 24 trials) and asks children to sort by color or shape. According to investigators 3 year olds are typically unable to pass the task, showing a pattern of inflexible responding when given instructions to switch and sort by a different feature. In contrast children of 5 years can usually complete the switch and by 7 years of age should be able to complete the final discrimination phase. In the DCCS instructions are given at each trial to constrain the interpretation and ensure it is not a failure of memory or testing other possible responses (See Table A5 for details on cards, targets and instructions). All discussion here refers to the standard version of the test. According to investigators most children over 36 months can sort the first six pre-switch trials correctly. The next six post switch trials are usually distributed such that a participant gets all the trials correct or all wrong.

Three primary questions were raised in this investigation. First, Forward and Backward Chaining IRA tasks were explicitly compared. The second, if a relationship would be detected between a measure of perseveration (the DCCS) and the IRA task. Third, if the relationship detected by Paule (1999) between backward chaining IRA accuracy and Full Scale IQ would be detected again and if this relationship extended to Forward Chaining IRA.

Chapter 2 Method

Participants

Children between the ages of 2.5 and 7 years of age were recruited from various day care centers and the general public. Data were collected over a four year period at six different locations; see Table A6 for details. All participants who were identified as having normal, or corrected to normal, hearing and vision (determined via parent report) were included in this study. Participants and their families received $10 in compensation for their participation.

A total of 65 participants signed up for this study. Due to scheduling difficulties, only 55 participants could be included in the study. Four of these participants did not complete the WPPSI-III but completed all other portions of the study, another four participants completed the WPPSI-III but were older than 7 years 3 months so their WPPSI-III scores were omitted from analysis. Data for these eight participants were used when possible. Sample size is noted for each individual analysis.

The final sample consisted of 29 male and 26 female children whose ages ranged from 32 to 94 months ($\overline{X}$ = 59.4 months): 1 child under the age of 3 years, 13 children between 3 and 4 years, 18 children between 4 and 5 years, 10 children between 5 and 6 years, 8 children between 6 and 7 years, and 5 children over 7 years old. During recruitment the upper age range specified was 7 years and any child under 8 years was allowed to participate.

Details of participants' demographic information is available in Table A7, and is based on parental self-report to a brief survey included with the information packet. A third of participant parents reported they had earned a college degree, another 38.18% of parents reported that they had completed high school and some college courses, 17.27% of parents reported a post graduate degree, other families left the question blank or indicated trade schools or an education that stopped prior to high school. Approximately half the sample came from families with an income between $25,000 and $74,000, and were predominately Caucasian (63.63%). Table A7 also includes information about TV viewing habits and Video Game playing habits. A high number of children reported having no (n = 21, no video game exposure) or very little (n = 28 less than 1 hour per day) experience with video games; despite this, it was clear that all of the children who participated in this study were well prepared to complete tasks on a touch screen computer. Because the question didn't specify details about games played on other platforms like tablets and computers it's not clear that this question truly captured the extent to which participants were prepared to work on our touch screen computer. Finally, participants who were reported to have a learning disability, such as Autism Spectrum Disorder (ASD), were not excluded from the study based (four participants were reported to have been diagnosed with ASD or ADHD). Data from these participants were included unless analysis indicated the participant was an outlier in terms of response total, data screening criteria are discussed in the Results section.

Prior to inclusion in the study a parent or guardian completed an informed consent for the participant. Each participant provided assent at the start of each research session. The study was approved by the Auburn University Internal Review Board and was conducted in accordance with all relevant guidelines.

Materials

WPPSI-III administration. The WPPSI-III is a measure of cognitive abilities in preschool age and early school age children. The test requires verbal and nonverbal responses to various stimuli, including blocks, puzzles, pictures and verbal questions (Wechsler 2002a). Administration requires a score sheet and specific blocks, puzzles and pictures. In this case the administration corresponded with the instruction manual requirements as closely as possible, occasionally it was necessary to break administration up over two sessions in order to facilitate the schedule of participants. Time to complete administration of the core subtests varies amongst individuals; approximately 30 minutes for children between 2:6 to 3:11, and approximately 1 hour for children 4:0 to 7:3. The WPPSI-III has been examined for reliability (coefficients for the various subtests, age groups above 0.8) and validity (intercorrelations were as expected) (Wechsler, 2002b). Participants ranging in age from 2:6 to 3:11 completed four subtests; Receptive Vocabulary (measuring the ability to comprehend verbal directions), Block Design (measuring analysis and reproduction of abstract design), Information (measuring recall of facts already known) and Object Assembly (measuring visual-perceptual organization, integration and synthesis of part-whole relationships, non-verbal reasoning, and trial-and-error learning). For participants in the 4:0-7:3 age band the seven core subtests were administered; Block Design (measuring analysis and reproduction of abstract design), Information (measuring recall of facts already known), Matrix Reasoning (measuring verbal reasoning, verbal comprehension, and general reasoning ability), Vocabulary (measuring knowledge of and the ability to express the meaning of words), Picture (measuring abstract, categorical reasoning ability), Word Reasoning (measuring verbal comprehension and reasoning) and Vocabulary (measuring knowledge of and the ability to express the meaning of words).

For this study the WPPSI-III was initially administered by graduate students in the clinical psychology program at Auburn University or by a graduate student enrolled in the University of Alaska Anchorage Clinical Community Psychology program. Later in data collection it was necessary for the investigator to collect all data without support from graduate students. Some test administrations, nine, were recorded and reviewed by the supervising committee for procedural integrity. When the WPPSI-III was administered by a graduate student that individual was blind to the child's IRA performance. The investigator was present during the administration of the WPPSI-III by graduate students and couldn't be completely blinded to the results of the either task. When it became necessary for the investigator to conduct the WPPSI-III testing without support the IRA testing occurred first whenever possible. When this was not possible the WPPSI-III results were not calculated until after all IRA testing was complete. Thus, although the investigator had a global observation of how each child was doing she didn't know the exact value of a child's WPPSI-III test until she could no longer influence their IRA results.

Dimensional change card sort task. The Dimensional Change Card Sort Task (DCCS) was used to measure a child's ability to switch tasks (Table A5) (Zelazo, 2006). This is similar to a Wisconsin Card Sort task, but has been designed for children of this age group. Children are asked to sort six cards by shape (Rabbit or Boat) and then six cards by color (Blue or Red). In a third phase (12 cards) the child must switch back and forth between these two attributes based on the presence or absence of a border. In previous studies participants were rated as simply passing/failing the procedure at each phase ("DCCS -phase" in this manuscript – phase 1 first phase, pass with 6/6 cards, phase 2 second set, pass with 5/6 cards, phase 3 final phase, pass with 9/12 cards). The order of cards presented to the participant and the instructions are presented in table form (Table A5, Zelazo, 2006). If the participant did not achieve 100% on phase 1 or 2 the procedure stopped. In order to ensure uniformity in this study all participants sorted all 24 cards and received the same instructions. Despite this change the use of total correct responses was inappropriate due to the high likelihood of obtaining correct responses via biased responding based on a side or pattern. In addition to phase a unique variable was calculated to provide the most information about the task possible. This variable "DCCS -correct" is not the *total* number of correct responses but the number of correct responses based on the phase criteria; for a participant who didn't master phase 1 correct responses were calculated based only on the first phase. When a participant mastered phase 1, but not phase 2, correct responses were calculated based on correct responses in phase 1 and 2. Finally for a participant who met the criteria for both phase 1 and phase 2, correct responses were calculated from all three phases. This allowed investigators to better discriminate between participants who passed phase 1 but not phase 2 and so on.

Computer hardware and software. The IRA task was conducted on a desktop iMac computer running Mac OS X Version 10.4.11 with an add-on Magic Touch™ (touch screen adaptor) device or a lap top iMac running the same system and fitted with an add on Magic Touch. The IRA program was developed using Realbasic (2007) programming by an undergraduate student programmer at Auburn University and was tested through several iterations by the investigators and then with college students prior to administration with children. Additional pilot testing occurred with a 3-year-old volunteer and a 7-year-old volunteer (approximate age). These were children of Auburn University professors who attempted the IRA task with their parents present; the data from these volunteers is not presented here or elsewhere.

Procedure

IRA task. During each IRA task the participant saw four squares with different colored triangles in them (Forward Chaining IRA) or four squares with different colored circles (Backward Chaining IRA). In all sessions a 9 link chain was the maximum chain length possible. All sessions started with instructions (italicized portions were modified slightly as appropriate for the exact research condition and stimuli used), "This is a guessing game. Your job is to find *Barney* hiding behind one of these buttons [Gesture to each button]." The investigator then pushed a correct button and said "Look, there he is!" Next the investigator pushed an incorrect button and said, "See, he's not here. It will get harder to find him the longer we play. You may have to push more than one button to make him show up. Try all by yourself right now, *later I will help you*. Try your best!" [Instructions: 64 words Flesch-Kincaid reading level 1.1 as determined by Microsoft word. The participants didn't read these instructions, however the reading level is provided for context.]

For all IRA tasks the computer provided feedback after every response chain; for a correct response chain the words "Good Job!" appeared on a green screen with one of several popular cartoon characters (Barney, Dora the Explorer, Bob the Builder, Thomas the Tank Engine). Incorrect response chains were followed by a black screen for 1 second. During all IRA sets the investigator sat next to the participant and verbally reiterated the consequence the computer displayed; if the response was correct the experimenter said "Good job!" etc. If the response was not correct the experimenter said "That's not it. Try again!" etc. An incorrect response reset the task to the beginning of the current chain. If the participant stopped responding or talked about other things the experimenter remind the participant to "Keep trying".

Occasionally a participant would ask to stop playing the game. The first time this occurred the participant was encouraged to "Try for a little longer." If the participant continued to ask to stop they were allowed to; this repeated request was treated as rescinding assent for that session. This almost never occurred over the course of the study four children asked to stop once, but only one persisted in stopping when encouraged to try. Data from these sessions were included unless analysis indicated the participant was an outlier in terms of response total, data screening criteria are discussed in the Results section.

This IRA task used brief audible tones as intervening stimuli, the arrangement of these stimuli during the IRA task is very important and although previously discussed is briefly described here. In addition Figure 1 is included for reference. In Figure 1, a 3 link chain is depicted; the numbers reflect a specific response and the letters represent a specific audible tone. It is necessary to note that even when the response sequence changed (as in a learning phase) the order of the audible stimuli did not change. During forward chaining the participant heard a stimulus at the start of each trial, tone A in Figure 1. When the chain incremented the participant heard the same initial stimulus, tone A in Figure 1, once they responded correctly they heard a second stimulus, tone B from Figure 1. In contrast during backward chaining the participant heard tone C from Figure 1 at the start of each trial until the chain incremented. When the chain incremented, the participant heard a new stimulus, Tone B from Figure 1, when this stimulus was followed by a correct response Tone C was encountered again.

Training sessions and error correction procedures. This training session differed from the other IRA tasks sessions in several ways. Only 3 response keys were available to make up a 9 link response chain. For the Forward Chaining IRA each response key contained a picture of a different piece of fruit. For the Backward Chaining IRA task each response key contained a picture of a different animal. In the absence of the error correction procedure the participants interacted with the computer task and the investigator in the same way as previously described for the initial session. The investigator reiterated the consequences displayed by the computer and reminded the participant to continue to engage in the task when necessary.

The participants were placed in one of three groups based on when they enrolled in the study, these groups are also depicted in Figure 2. Participants in Group 1 (automatic error correction, described below), the first participants to enroll, automatically received an error correction procedure from the research during the “Training Session” each day that the IRA task was encountered. Participants in Group 2 (contingent error correction), who enrolled later in the study, received an error correction procedure from the investigator only on the second day of the IRA task and only if they did not successfully emit a 2 link chain on the first research day. Participants in Group 3 (2 days only, no error correction), the last to enroll in the study, did not receive any error correction procedure as they worked on each IRA task on only one day. Even though participants in Group 3 didn't encounter the error correction procedure they, like participants in Group 2, experienced the training session as a second task between the two primary IRA sessions but without intervention from the investigator.

The error correction procedure meant the investigator interacted more with the participant. This procedure was initially intended to modify repetitive errors, not prompt the participant to a specific correct response. This was done by labeling contingencies and significant stimuli and feedback provided by the program. Gestures and physical prompting were only used as needed throughout the program. The specific protocol for this is displayed in Table A8. During data collection it became clear that many errors were a result of participant swiping at the touch screen (like one would with a tablet or smart phone), or responding too quickly for the computer to read a response (pushing all four buttons before the computer could display the reinforcer screen or the blackout screen). These problems were addressed during the training session for all participants, regardless of the experimental group they had been placed in. This was addressed by giving the instructions "Don't swipe just touch the button." and "Slow down, the computer can't go that fast." each time the participant displayed one of these behaviors during the training sessions.

Task presentation. Participants worked with the investigator on up to 5 different days to complete the Wechsler Preschool and Primary Scale of Intelligence-III, the Dimensional Change Card Sort Task, Forward Chaining IRA and Backward Chaining IRA. The order in which participants experienced the various elements of the study was flexible; based on the scheduling needs of the investigator and the participants some participants completed all IRA tasks prior to starting the WPPSI-III, others completed the WPPSI-III prior to starting the IRA task. The IRA sessions were divided so that two research days in a row were Forward Chaining IRA task and two research days in a row were Backward Chaining IRA task. The order in which participants experienced the different IRA tasks was counterbalanced across participants. Near the end of data collection a decision was made to shorten the study in order to ensure the study was completed in a timely manner. For some participants only three research days was necessary, the participants only experienced the IRA task on two occasions; one Forward Chaining IRA task and one Backward Chaining IRA task. The various IRA task sessions and days are depicted in Figure 2. All participants' first encounter with the first session of each IRA task was identical but subsequent sessions varied based on the experimental group. Each IRA session lasted 5 minutes. Participants were given the opportunity to complete up to three IRA sessions on each research day that included the IRA task.

Data Analysis

Three primary questions requiring statistical analysis were proposed in this investigation. First, if the relationship detected by Paule (1999) between backward chaining IRA accuracy and Full Scale IQ would be detected again and if this relationship extended to Forward Chaining IRA. The second, if a relationship would be detected between a measure of perseveration (the DCCS) and the IRA task such that the measure of perseveration was associated differentially with performance on forward and backward chaining. Third, forward and backward chaining tasks were explicitly compared. Finally, several secondary questions developed based on determining how an error correction procedure affected IRA performance, the impact of the order in which IRA tasks were experienced and the extent of participant improvement over sessions. The analyses used to consider these questions are discussed in this section.

Three primary dependent measures of the IRA task were used to ensure a comprehensive perspective was gained from the data: response total, chain length and PQ. Accuracy and error totals were also available and used when they provided additional information about a specific question. The WPPSI-III Full Scale IQ and two composite scores, Verbal IQ and Performance IQ, were included. Two measures from the DCCS were used: DCCS-phase and DCCS-correct.

Prior to analyses each of the dependent measures from the IRA, WPPSI-III and DCCS were examined in relation to participant's socio demographic characteristics: gender, race, income, language use and exposure and reported learning disabilities or special service. These examinations included research site and error correction groups. Separate one-way analyses of variance (ANOVA) were conducted for each measure with the various socio demographic variables serving as between subject factors in order to detect any unexpected relationships with our measures of interests.

Two additional variables might affect performance on the IRA task and became the subject of secondary analyses, "Order of IRA Task" presentation and "Error Correction Group." In addition a comparison between the IRA Tasks and the four sessions is made to determine if participants improved after the first encounter with the IRA task. A first set of repeated measure analyses of variance (RMANOVA) included all of these factors and were conducted for total response, chain length and PQ. These RMANOVAs included two between-subject factors, order of IRA Task Experience and Error Correction Group, as well as two within-subject factors, IRA Task (Forward and Backward) and Session (1,2, 3 and Final). For this and all other RMANOVAs a Bonferroni correction was applied when planned comparisons or post hoc comparisons were necessary.

The main effects of Error Correction Group and Session were examined in detail through a series of RMANOVAs with Error Correction Group as the first factor (between-subject) and Session and IRA Task as additional factors (within-subject). These ANOVAs were conducted separately for the two IRA chain types and for each IRA measure. The use of multiple RMANOVA does make Type 1 error a risk, but these additional tests were used for exploration not hypothesis testing of our primary research questions.

To further develop the effect of Session in relation to IRA Task the Difference in performance between Session 1 and Session 3 (See Equation 2) was calculated and was used in another set of RMANOVAs. In these RMANOVAs IRA Task and the Order of IRA experience were the only factors entered.

Equation 2: Difference Value (2)

Difference Value = (Value at End – Value at Beginning)

The main effect of the Order of IRA Task presentation was further examined with a set of RMANOVAs calculated for each of the IRA measures. For these analyses only one between-subject factor was entered, Order of IRA Experience and one within-subject factor was entered, IRA Task.

The associations between IRA Session 1 (combined with Session 5), Session 3 (combined with Session 7) and IRA Difference Values (Session 1 to Session 3) with participant age, IQ and DCCS were examined using a series of Pearson correlations. IQ was used to separate participants into 3 groups for an ANOVA conducted separately for Forward Chaining IRA Session 1 and Backward Chaining IRA Session 1. A regression was conducted separately for Forward Chaining IRA and Backward Chaining IRA Session 1 data with IQ and DCCS as predictors.

Chapter 3 Results

Comparing Two IRA Tasks

The exploratory analyses are presented first, before proceeding to the primary questions proposed in this study. Due to the relative novelty of the Forward Chaining IRA task it is necessary to examine how the two tasks compare with each other and how they relate to the procedural manipulations controlled within this study: error correction strategy, order of the IRA task presentation (forward or backward chaining first) and exposure to multiple sessions. To form these comparisons, data from the 29 participants who experienced each IRA task over 4 sessions and who passed data-quality screening (immediately below) were used. See Figure 2 for additional details on the participants' experiences.

Pre-screening of data. Prior to analysis of any IRA session the total number of responses emitted was examined for outliers and participants with an abnormally low number of responses. First each IRA task and session was screened using box plots for outliers. Extreme outliers, more than three standard deviations below the mean, were removed from further analyses. Remaining cases in which participants responded fewer than 10 times during any session were then examined in greater detail. If these participants also never advanced to a two-link chain they were removed from further analysis. No participants were removed for responding at a high rate. Together, the screening resulted in the removal of two sessions from two different participants, one with 18 and one with 3 responses.

Description of all IRA task manipulations. The change in response total, chain length and PQ scores for participants who completed four sessions are displayed in Figure 3. This figure, and subsequent analyses, makes use of data from 29 of the 55 participants who completed all eight IRA sessions. Participants were grouped based on the order in which IRA tasks were experienced and the Error Correction Group participants were placed in. This sub-sample was not randomly selected for placement within these groups (group placement was based on date of enrollment and is described in detail in the methods section). Table A9 in the appendix displays the mean and standard deviation of the IRA measures shown for each group within Figure 3.

Participant age and IQ were used to characterize differences among the four groups formed (Table A10). There were no statistical differences between these four groups in either participant age or IQ, all *p*'s > 0.1 (from univariate ANOVAs not displayed). The mean age of participants was almost a year lower for the Forward Chaining First, Automatic Error Correction participants than other groups. IQ ranged among these four groups from a mean of 101 to a mean of 115 (note that for the group with a high mean IQ of 115 only two participants completed IQ testing).

A direct comparison of the Forward and Backward IRA tasks was conducted using a series of RMANOVAs. This analysis is displayed in Table A11 (which shows all relevant values: df, SS, MS) and a truncated version of this table, showing only F and *p* values, is displayed in Table 1 This analysis contained two within-subject factors (Session and IRA Task) and two between-subject factors (Error Correction Group and the Order of IRA Experience). The results of these analyses are examined in this section and then broken down further to examine the difference between the Forward and Backward Chaining IRA task.

On the Backward Chaining IRA task participants generally emitted the most responses on their first encounter with the IRA task. On the Forward Chaining IRA task response totals were unstable across sessions, but were generally lower in later sessions within an IRA Task, $F(3,57) = 3.98$ $p = .012$. There were fewer responses on the Backward Chaining IRA task compared to the Forward Chaining IRA task, $F(3,57) = 3.19$ $p = .030$.

Chain length increased across IRA sessions 1-3 and 5-7 but then decreased in the final "learning" session (session 4 or 8), when a novel chain was presented $F(3,57) = 4.51$ $p = .007$. Chain length was higher on the second IRA task encountered. The Error Correction Groups did not differ on chain length in a consistent direction (i.e. one Error Correction group did not consistently out-perform the other). A modest improvement in PQ occurred across IRA sessions 1-3 and 5-7 but performance didn't remain high in the final "learning" session, $F(3,57) = 6.85$ $p = .001$. Unlike chain length, PQ did not improve on the second task encountered.

One notable result was the absence of difference, statistically speaking, between forward and backward chaining when the participants were separated based on Error Correction Group and Order of IRA Experience: there was no main effect of chain type. In order to examine the other effects of interest (Error Correction Group, Order of IRA Experience, and Session) with more power the two tasks were examined individually in some of the additional analyses.

Error correction procedure. The analysis of the error correction procedures focuses on determining if the approach affected performance and if the two groups can be combined. Visual assessment of the right two panels of Figure 3 suggests that the error correction procedure was more effective for Backward Chaining IRA than Forward Chaining IRA, especially on chain length and PQ, but this apparent difference between the two tasks did not rise to statistically significant levels when the two tasks were compared directly. When the IRA tasks were considered separately Error Correction Group was still not a significant main effect for either of the IRA tasks.

Although there was no main effect of Error Correction on total responses, chain length or PQ in the first set of RMANOVAs (Table 1 and Table A11), there were interactions with the Order of IRA Experience that suggested further scrutiny was appropriate. [Response total $F(1, 19) = 9.08$ $p = 0.007$, chain length $F(1, 19) = 4.46$ $p = 0.048$ and PQ $F(1, 19) = 5.32$ $p = 0.033$.] When RMANOVAs were conducted for the two IRA tasks separately (Table A12) most of these interactions disappeared. The exception was Forward Chaining IRA response total where the Error Correction Group interacted significantly with the Order of the IRA Task, $F(3, 63) = 4.37$ $p = 0.049$.

Regardless of whether the Forward Chaining IRA task occurred first or second, there appeared to be no differential impact of the Error Correction Procedure on measures of performance (chain length and PQ). The same was true for the Backward Chaining IRA task. Regardless of when the Backward Chaining IRA Task occurred, first or second, there was no statistical difference between the error correction procedures.

There was, however, an interaction between Error Correction Group and Session on chain length for the Backward Chaining IRA Task, $F(3,69) = 3.51$ $p = 0.020$. For the Contingent Error Correction Group, performance improved (in terms of chain length and PQ) on the session immediately subsequent to the error correction procedure. For the Automatic Error Correction group no improvement was apparent after the first error correction procedure in terms of chain length. Modest improvement was apparent on PQ, but was not statistically significant. By Session 3 the two Error Correction Groups' performances were almost indistinguishable. Visually the Error Correction Groups did not follow the same pattern when Backward Chaining IRA occurs first or second, but as previously noted these differences did not rise to statistically significant levels when only Backward Chaining IRA is considered.

There was, at best, a minimal impact from an error correction procedure and the timing of an error-correction procedure; so to increase the power of further analysis, data from the two Error Correction Groups were combined. This was accomplished by selecting data from session 1, 3, 5 and 7, sessions that marked points when both groups had experienced none (session 1 and 5) or one (session 3 and 7) error correction procedure. For Session 3, 5 and 7 on both IRA tasks the differences between the two Error Correction groups were quite small (based on visual assessment) and were not statistically significant (one-way ANOVAs not displayed, all p's > 0.2). This was also true for Session 1 on the Forward Chaining IRA task (p's > 0.1) but not for Session 1 on the Backward Chaining IRA task where the two Error Correction Groups differed on the chain length attained ($F(1,11) = 9.44$ $p = 0.012$). Still; in subsequent analyses of Session 1 3, 5 & 7 the Error Correction groups were combined.

Performance in multiple IRA sessions. There was a significant main effect of session in the first set of RMANOVAs calculated (Table 1 and Table A11) for response total, chain length and PQ. Performance on the IRA Tasks generally improved over sessions. Session also interacted with Error Correction Group and Order of Experience, but these interactions were for response total only, not chain length or PQ. When the two IRA tasks were examined separately (Table A12) the pattern of significant interactions with Session was minimized; session affected only response total on the Forward Chaining IRA task (Fig 4, top). As previously noted, although somewhat erratic, the response total showed a downward trend over sessions, $F(3,63) = 3.95$ $p = 0.012$, and was lower on the Backward Chaining IRA Task, $F(3, 63) = 4.45$ $p = 0.007$.

Although the two IRA tasks did not statistically differ from each other in a direct comparison (Table 1, or Table A11 and Figure 4), when analyzed separately, the improvement in performance was not the same for the two IRA tasks. For Forward Chaining IRA, Session remained a significant main effect on chain length [F(3,63) = 4.23 p = 0.009] and PQ [F(3,63) = 5.76 p = 0.002] both improved across the three performance sessions. In contrast, on the Backward Chaining IRA Task Session was not a significant main effect for response total or chain length. PQ did improve across the three performance sessions but didn't remain high in the final "learning" session (Figure 4, middle and bottom), F (3, 69) = 2.80 p = 0.046).

The Difference between Session 1 and 3 and between Session 5 and 7 were calculated (See Formula 2 in the Data Analysis section). This formula yielded negative values for some individuals, which represent a decline in performance between Session 1 and 3, or between Session 5 and 7. When the participants are considered as a group a decline is only visible in the response total between Session 5 and 7. The means and standard deviations of the Difference Values are displayed in Table A13. For all measures there was wide variability in the change in performance between the sessions as evidenced by the extreme range and high standard deviations.

A series of RMANOVAs on the Difference value with the Order of IRA Experience as a between-subject factor and IRA Task as a within-subject factor is displayed in Table A14. The two tasks differed in terms of change in response total [F(1,23) = 6.96 p = 0.015] chain length [F (1,23) = 8.67 p=0.007.] and PQ [F(1,23) = 4.92 p = 0.037]. Response rate increased more relative to the initial value for Forward Chaining IRA than Backward Chaining IRA. Chain length also increased more between the sessions for Forward Chaining IRA. The Order of IRA Experience was not a statistically significant main effect on the change in any of the dependent measures. There was also no significant interaction between the Order of the IRA Experience and the IRA Task on the change in any of the dependent measures.

Effect of the order in which IRA chain types were experienced. Participants were semi-randomly assigned (as described in the methods section) to experience either Forward or Backward

Chaining IRA first. Order of IRA Experience interacted with the Error Correction Groups (discussed in the Error Correction section) and IRA Task on chain length $F(1,19) = 4.74$ $p = 0.042$. The interactions between Order of IRA Task and Error Correction Group were explored in the previous section. An investigation of the interaction between Order of IRA Task and IRA Task follows.

To further examine the interaction between Order of IRA Task and the IRA Task a new set of RMANOVAs was conducted for the selected sessions only (1, 3, 5 & 7) and with the Error Correction Groups combined. The results are displayed in Table 2. There was a main effect of order of IRA Experience on PQ [$F(1, 23) = 1.20$ $p = 0.049$], but not on response total or chain length. In Figure 4 it is clear that the group that started on the Backward Chaining IRA Task had lower PQ scores on Sessions 1 and 3, but higher PQ scores on Sessions 5 and 7, during which they experienced Forward Chaining. In this analysis the IRA Tasks did differ from each other to statistically significant levels for all three IRA measures: response total $F(1, 23) = 13.22$ $p = 0.001$, chain length $F(1, 23) = 5.71$ $p = 0.025$ and PQ $F(1, 23) = 5.22$ $p = 0.032$. Visual inspection of Figure 4 makes clear that for the sessions selected (1, 3, 5, & 7) when Error Correction Groups were combined, it was not the order of the IRA tasks which matters most but the IRA task itself. For both of the performance measures (chain length and PQ) individuals performed better on the Forward Chaining IRA task regardless of what order they experienced the tasks in. This is different from the earlier results when all independent variables and sessions were included, and IRA chain type did not appear to differentially impact the results.

Comparison of forward and backward chaining IRA. When only sessions 1, 3, 5 and 7 were considered and the Error Correction Groups were combined, participants consistently performed better on the Forward Chaining IRA Task, whether it came first or second. In addition participants improved more between the selected sessions on the Forward Chaining IRA Task than the Backward

Chaining IRA Task. The error correction procedures were more effective for the Backward Chaining IRA Task than the Forward Chaining IRA Task but this difference did not rise to statistically detectable levels.

The differences in performance measures between the two IRA tasks could be attributed the widely different (statistically significant) response totals emitted by participants on the two tasks themselves. Analyses thus far have not made use of raw measure of IRA performance, such as error total. In Figure 5 the error total is juxtaposed with the response totals for each session. There were no statistical differences between the two IRA tasks on Error Total (from RMANOVA not displayed). For each IRA task there was a decrease in errors between the selected sessions. For the Forward Chaining IRA task there is a corresponding increase in responses. In contrast for Backward Chaining IRA responses decrease. Inspection of Figure 4 also shows that response total did not increase between sessions on the Forward Chaining IRA task the same way when forward chaining was encountered second, this difference didn't rise to statistically significant levels and is probably a reflection of uncontrolled group differences. This difference didn't affect the increase in response total seen in Figure 5 when all participants were combined.

To further examine these differences an RMANOVA of percent errors was conducted, and didn't reveal any new information in regard to the error correction procedure and the order of IRA task. When these groups were omitted from analysis the two IRA tasks were not statistically different from each other; only session remained a significant variable with participants having a significantly lower error percentage on Forward Chaining IRA session 3 (43% vs > 53% on session 1 and both sessions of the Backward Chaining IRA task) $F(1,24) = 30.44$ $p = 0.000$.

Correlation between measures of Forward and Backward Chaining IRA. Analyses thus far have focused on the difference between the two IRA Tasks but not looked for similarities between the two tasks. To determine if an individual's performance on Forward and Backward Chaining IRA were related, inter-correlations were calculated (Table A15). These calculations included five measures of the IRA task: response total, error total, chain length, accuracy, PQ. The Difference Values for response total, chain length and PQ were also correlated. In these calculations participants in Group 3 (truncated exposure to IRA task) were included whenever possible. The error correction groups were combined and the sample was not divided based on the order the IRA tasks were experienced in. There were no statistically significant correlations among any of the IRA measures for the session 1 and 5 (all r's < 0.26). The two IRA tasks only correlated on one of the measures for Difference Values, response total $r = 0.539$ $p = 0.005$.

Association between IRA tasks and age, IQ and DCCS scores

Description of WPPSI-III. Forty-six participants completed the IQ testing, the mean Full Scale Composite score for the entire sample was $\overline{X}$=103.22 SD = 12.43 (range 78-133). The Verbal Composite for the sample was $\overline{X}$ =103.26 SD = 13.285, and the Performance Composite for the sample was $\overline{X}$=100.43 SD = 13.55. A description of the WPPSI-III scores is located in Table A16. Several one-way ANOVAs were conducted in order to examine the WPPSI-III scores and the demographic characteristics of participants. These analyses included gender, race, income, language use and exposure, learning disabilities or special service reported, research site and error correction group. No significant effects on WPPSI-III scores were detected based on race, income, language use and exposure, or learning disabilities or special services reported. An effect did exist for gender, research site and error correction group on the Verbal IQ. These differences were also displayed in Table A16.

Females obtained higher Verbal IQ scores than males, $F(1, 45) = 5.040$ $p = 0.03$. Research location also was related to Verbal IQ, $F(5, 45) = 2.854$ $p = 0.027$. Post hoc testing revealed children at Location B (n = 17) outperformed those at Location C (n = 13) on the Verbal IQ. This may be related to the different participant make-up of the two locations. A review of the demographic information shows that the sample from Location C was more likely to be male (9 of 13 participants) and come from a lower income background (11 of 13 participants reported an family income less than $50,000) compared to Location B (5 of 17 participants were male, 6 of 17 reported a family income less than $50,000).

Dimensional Change Card Sort task. Fifty three participants completed the DCCS task. For DCCS-correct responses (based on phase criteria) there was mean of 12 and a response range of 0-23. DCCS-phase was a categorical variable; 6% of participants did not pass the first phase, 32% of participants passed the first phase, 30% of participants passed the first two phases. Descriptions of the DCCS measures are located in Table A17.

The DCCS measures were examined using the same demographic variables as the WPPSI-III: gender, race, income, language use and exposure, learning disabilities or special service reported, research site and error correction group. There were no significant effects for any of the participant characteristics examined, these ANOVAs were not displayed.

Correlations between participant age and the DCCS were expected and found. For DCCS-phase, the correlation with age was $r\ (53) = 0.523$ $p = 0.000$ and DCCS-correct also positively correlated with age $r\ (53) = 0.534$ $p < 0.000$. The DCCS was also compared to the WPPSI-III. For all three WPPSI-III components there were no significant correlations with DCCS measures. This lack of correlations was true regardless of the DCCS measure used, these correlation values are available in Table A17.

IRA task. The investigation of the associations between the IRA Tasks and age, IQ and DCCS scores generally focused on data from sessions 1 and 5 (referred to together in this section as

Session 1). Sessions 3 and 7 (referred to together in this section as Session 3) were also examined. These data were not divided based on error correction group or the order the IRA Tasks were experienced, the two error correction groups were combined, and data from Session 1 and 5 (or Session 3 and 7) was combined for each IRA task. The mean, SD and Range of each IRA task is presented in Table A18. Like the WPPSI-III and the DCCS, IRA measures were subjected to exploration via means testing on several participant characteristics: gender, race, income, language use and exposure, learning disabilities or special service reported, research site and error correction group.

For the Forward Chaining IRA Task (n=52) there were no significant effects for any of the participant characteristics examined for any of the Forward Chaining IRA measures (ANOVAs not reported). For Backward Chaining IRA (n=53), an effect was found for the language exposure characteristic. The descriptive information for this relationship is also displayed in Table A18. A participant's language exposure related to two measures of Backward Chaining IRA, accuracy $F(2, 50) = 3.866$ $p = 0.027$ and PQ $F(2, 50) = 4.438$ $p = 0.017$. Post hoc analysis indicated that for accuracy and PQ participants with exposure to Spanish in the home or who were reported to be fluent in Spanish out-performed those who spoke only English.

Forward chaining IRA task Session 1. Correlations were calculated among five measures of the IRA Task and participant age, the DCCS and the WPPSI-III composite scores. These correlations were displayed in Table 3 and were graphically depicted in Figure 6. Age was positively and significantly correlated with the number of responses emitted [r (52) = 0.279 p = 0.046] and the PQ score, r (52) = 0.275 p = 0.049. Both measures of the DCCS positively correlated with the number of responses emitted [r (51) = 0.362 p = 0.009, r (51) = 0.338 p = 0.015]. Two of the WPPSI-III composite measures, Full Scale IQ and Performance IQ, were positively correlated with the number of errors committed: r (44) = 0.313 p = 0.039, r (44) = 0.391 p = 0.008 respectively. Several non-significant negative correlations were noted between Verbal IQ and the various measures of the IRA task: response total r (44) = -0.177, chain length r (44) = -0.101 and accuracy r (44) = -0.264.

Correlations were also calculated among the raw WIPPSI-III subtests and the Forward Chaining IRA measures. These correlations were run separately for the two WIPPSI age groups; 2:6-3:11 and 4:00-7:3. There were no significant correlations with the measures of IRA performance, chain length and PQ, for either of the age groups. For the younger age group response total and error total did correlate with Block Design [r (9) = 0.724 p = 0.028, r (9) = 0.727 p = 0.027], Information [r (9) = 0.729 p = 0.026, r (9) = 0.793 p = 0.011], Object Assembly [r (9) = 0.711 p = 0.032, r (9) = 0.721 p = 0.028].

Backward chaining IRA task Session 1. Correlations were calculated between the five measures of the IRA Task and participant age, the DCCS and the WPPSI-III composite scores. These correlations were displayed in Table 4 and are graphically depicted in Figure 7. Age was correlated with the number of responses emitted, r (53) = 0.510 p = 0.000. Neither of the DCCS measures correlated with any of the IRA task measures. Both Full Scale IQ and Performance IQ positively correlated with the error total: r (45) = 0.362 p = 0.013, r (45) = 0.378 p = 0.010 respectively. In addition Performance IQ also correlated with the number of responses emitted, r (45) = 0.365 p = 0.013.

Correlations were also calculated among the raw WIPPSI-III subtests and the Backward Chaining IRA measures. These correlations were run separately for the two WIPPSI age groups; 2:6-3:11 and 4:00-7:3. For the younger group correlations only existed between response total and Block Design [r (10) = 0.655 p = 0.04], Information [r (10) = 0.639 p = 0.047], and Object Assembly [r (10) = 0.844 p = 0.002]. Among the older group there was one significant correlations with a measures of

IRA performance, PQ and the Vocabulary subtest were related, r (39) = 0.316 p = 0.05. There were also significant correlations between response total and Block Design, Information, Matrix Reasoning, Vocabulary, and Coding (all r's > 0.317).

Forward chaining IRA task Sessions 3. Correlations were calculated between the five measures of the IRA task and participant age, the DCCS and the WPPSI-III composite scores. These correlations are displayed in Table A19. Although these correlations come from a smaller sample than those calculated for Session 1, many of the correlations calculated were stronger.

For Session 3, participant age showed a positive correlation with all Forward Chaining IRA measures [n = 28, all r's > .37, all p's < 0.04] except error total. Full scale IQ and Performance IQ were not associated with any of the IRA measures. Verbal IQ showed several strong negative correlations with the Forward Chaining IRA session 3 measures: response total r (23) = -0.544 p = 0.007, chain length r (23) = -0.447 p = 0.032 and PQ [r (23) = -0.436 p = 0.038. Both of the DCCS measures were positively correlated with the number of responses emitted [r (28) = 0.400 p = 0.035, r (28) = 0.394 p = 0.038] and the IRA PQ: r (28) = 0.393 p = 0.039, r (28) = 0.422 p = 0.025.

The difference between Session 1 and 3 was also examined in relation to age, WPPSI-III and DCCS. Participant age was significantly correlated with the number of responses emitted [r (26) = 0.716 p = 0.000], chain length [r (26) = 0.601 p = 0.001] and PQ [r (28) = 0.540 p = 0.004]. The only other significant correlation existed between Verbal IQ and response total, r (21) = -0.537 p = 0.012.

Backward chaining IRA task Sessions 3. Correlations were calculated among the five measures of the IRA Task and participant age, the DCCS and the WPPSI-III composite scores. These correlations were displayed in Table A20.

For session 3 participant age was correlated with the number of responses emitted, r (29) = 0.643 p = 0.000. Verbal IQ was negatively correlated with the number of responses emitted, r (24) = -0.460 p = 0.024. Performance IQ was correlated with accuracy, r (24) = 0.529 p = 0.008. Both of the DCCS measures were correlated with the number of responses emitted: r (29) = 0.438 p = 0.018, r (29) = 0.379 p = 0.043. In addition the DCCS Correct was correlated with the total number of errors, r (29) = 0.368 p = 0.049.

The difference between session 1 and 3 was also examined in relation to age, WPPSI-III and DCCS. Participant age was positively correlated with the number of responses emitted, r (29) = 0.490 p = 0.007. Negative correlations existed between Full Scale IQ and response total [r 924) = -0.414 p = 0.044] and between Verbal IQ and response total [r (24) = -0.636 p = 0.001]. Both of the DCCS measures were correlated with the number of responses emitted: r (29) = 0.409 p = 0.028, r (29) = 0.378 p = 0.043.

Regression of Session 1 IRA. A linear regression was conducted separately for each of the IRA tasks. Two independent variables were used; Full Scale IQ and DCCS Phase. The dependent variable was IRA PQ from Session 1. For both chain types positive skewness was detected in the PQ: Forward Chaining IRA PQ had a skewness statistical value of 1.834 and a standard error of 0.330, Backward Chaining IRA PQ had a skewness value of 2.210 with a standard error of 0.327. The PQ measure for both chain types was subjected to a Log 10 transformation due to the lack of normality. In both cases this eliminated extreme outliers. The value for Forward Chaining IRA PQ skewness statistic was reduced to 1.206, standard error of 0.330. Backward Chaining IRA PQ had a skewness statistic of 1.220, standard error of 0.327.

The regression for a Log10 Forward Chaining IRA PQ was not significant, $F(2, 41) = 1.120$ $p = 0.336$. The regression for Log10 Backward Chaining IRA PQ was not significant, $F(2, 43) = 0.548$ $p = 0.582$. The values for B, Standard Error and β were included in Table A21.

Chapter 4 Discussion

The current work makes two novel contributions to the literature describing the IRA procedure in humans. The first is the comparison of two methods of chain development, forward and backward chaining. Second, this study represents the first attempt at examining a true Incremental Repeated Acquisition procedure with children, as opposed to an "incremental acquisition task" (Baldwin et al., 2012) or single encounter with the IRA task. Of equal importance, the study provides an expansion of previous work regarding the association between IQ, age and performance on a backward chaining IRA task. This addition comes from the inclusion of younger participants and a measure of executive function, the DCCS a card sort task, which can reveal perseverative tendencies.

Differences were detected between the two IRA tasks, forward and backward chaining. Participants performed better (longer chain lengths, higher PQ scores) on the Forward Chaining IRA Task than the Backward Chaining IRA Task. They also improved more on the Forward Chaining IRA Task between Sessions 1 and 3 (or Sessions 5 and 7). The Forward Chaining IRA Task also appears to have maintained a higher response rate based on the response totals. Superior performance on the Forward Chaining IRA task was consistent with previous (unpublished) studies with college students, but still a surprise given the literature doesn't support a strong expectation of a difference between the two tasks. The difference in performance on the two IRA tasks may be related to the different possible roles the discriminative stimuli play in Forward and Backward Chaining IRA, but this study was not designed to determine this difference specifically.

For the Forward Chaining IRA Task, performance improved across the three performance sessions and this improvement was observed for both Error Correction Groups (these two groups were almost indistinguishable on most measures of the Forward Chaining IRA Task). Performance on the Backward Chaining IRA Task appears to have been helped more by the error correction procedures, but this effect was subtle. Improvement on the Backward Chaining IRA task, although present in some measures, was less dramatic. In addition it appeared that when the error correction procedure was delayed and contingent on prior performance it had a greater impact on Backward Chaining IRA Task performance compared to an error correction procedure that occurred automatically and early in the investigation.

The inclusion of multiple IRA sessions did show that IRA performance improved after the first encounter, and this was true for both forward and backward chains. Although this implementation was novel, it is necessary to point out that the multiple measures occurred during what is called a “performance phase” where the same sequence was encountered during each session. This is still a departure from an IRA task as typically administered within most animal laboratories where an animal encounters a “learning phase”, a new sequence each session. Unfortunately the length of the study didn’t allow for repeating the IRA task until steady-state performance could be observed on the task, and the first encounter with the IRA task was still used to answer many of the research questions of interest within this study. As noted by Baldwin (2012) the logistical challenges of measuring the IRA task multiple times in humans are a real barrier to using the task more extensively with humans. In addition the retrospective inclusion of the error correction procedure limited the value of some of the IRA data that was collected.

Paule and colleagues (1999) first noted the correlation between a backward chaining IRA task and IQ. This relationship was further described in a study published by Baldwin (2012). This newer work confirmed that both age and IQ account for a child's ability to complete a backward chaining IRA task by comparing participants of various IQ levels, as opposed to using correlations. Taken together this work showed that, for children between 4 and 8 years, accuracy on the IRA task increases with age (Paule, Forrester et al., 1990b). In children 6 years of age there was a significant correlation between accuracy on the IRA task and full scale IQ scores (Paule, et al, 1999). Baldwin's work (2012) encompassed children between 5 and 13 years and evaluated variables beyond response rate and accuracy to provide a more nuanced picture of IRA performance. Specifically, Percent Task Complete increased with age and IQ; by age 8 years of age, high IQ participants were no longer distinguishable from participants within the average IQ group. Low IQ participants remained distinguishable until age 10. A similar pattern was found for Memory Response Accuracy with differences between the IQ "groups" disappearing between age 11 and 12. Response chain accuracy also increased with age and IQ, but the effects of IQ did not diminish with age.

The findings of the current study are more complex. The current study failed to find a statistically significant positive correlation between accuracy (or any other measure of performance such as chain length or PQ) and Full Scale IQ on either of the IRA tasks for data from Session 1 for participants on the lower end of the age range. The Session 1 data also yielded unanticipated positive correlations between IQ and the error total on both IRA tasks. Despite this there was a significant positive correlation between Session 1 Forward Chaining IRA PQ and participant age. For both tasks, age positively correlated with the Session 1 response total. Age and IQ appear to relate to performance on the IRA task in different ways. In this case, as age increased so did responses, but for children with a higher IQ error totals were elevated. It is possible that children with high IQ's "explored" the task more, attempting more novel response variations in an effort to obtain a correct response.

The Session 3 data were even less supportive of a relationship between IRA and IQ. For Session 3, Forward Chaining IRA there were significant negative correlations between Verbal IQ and response total, chain length and PQ. It is not clear why a high Verbal IQ wouldn't be helpful to performance of an IRA task, or why it might even be detrimental to performance of an IRA task. There was also a negative correlation between Verbal IQ and the number of responses emitted and Session 3 Backward Chaining IRA. In the Session 3 data we begin to see divergence between the two IRA Tasks, Forward Chaining IRA exhibited strong positive correlations with participant age. In the Backward Chaining IRA the response total was the only IRA measure that correlated with participant age. Again it is notable that for Session 3 the correlations between IQ and error total were positive, but for the difference between the IRA sessions there is a negative correlation between IQ and error total indicating high IQ participants reduced the number of errors more between sessions.

The data from this investigation only weakly emulated the expected connections between Backward Chaining IRA, IQ and age. Although the current study was never intended as a replication of Paule (1988b, 1990b, 1999) and Baldwin's (2012) work the differences between the investigations are worthy of consideration. Like previous work, performance on the first encounter with an IRA task was better for older participants, but neither of the IRA tasks correlated with IQ in a meaningful way in the age groups studied here. Given the apparent lack of correlations our discussion turns to procedural differences between the two IRA tasks that might help explain the different results.

The inclusion of children under 4 years of age in a study that included IQ as a variable represents a departure from earlier work on the IRA task, but earlier work indicated that IQ differences were most pronounced in younger children. Still, the relation of age on the various IRA measures was in the direction expected, although it was not always strong with the IRA measures this paper has deemed most descriptive of the task. The selection of measures appropriate to describe the IRA task can be difficult and can be affected by relatively minor procedural differences between various IRA tasks.

The IRA task used by Paule and his colleagues (1988b, 1990b, 1999) has a constrained number of reinforcers (3 at each level of a 6 link chain for a total of 18) and a less stringent advancement criterion between response chains. This contrasts sharply with the design of the IRA task in the current project where consecutive correct responses were required to advance through the chain and thus the number of reinforcers available throughout the task is unlimited. Paule and his colleagues (1988b, 1990b, 1999) used relatively few measures of IRA in early work: general accuracy, percent task complete (based on the # of reinforcers obtained which was constrained at each chain length) and response rate. In later work by the group Baldwin (2012) added various measures of accuracy based on different types of responses in order to produce a more comprehensive picture of responding within the IRA task and these were sensitive to the IQ groupings. As previously discussed overall accuracy is not a good choice for a direct comparison of the two IRA chain types when there is a mastery criterion, but it was calculated and used during analysis in this project, and was not correlated with either age or IQ for either chain type in this study. In the current work the PQ is favored (progress quotient developed by Bailey, Johnson and Newland in 2010) because it is more sensitive than accuracy to the differences in chain length, provided the participant progresses beyond a one link response chain. Because many participants in this study didn't exceed a one link response there appears to be a "floor effect" when PQ is examined; but a one response chain is the shortest chain possible and there is no restriction of range upwards making this apparent floor effect informative in regards to the difficulty of the task. This paper also used response total and error total as raw measures of the IRA task but because responses are unconstrained these measures must be interpreted carefully.

As noted earlier, high IQ participants actually incurred more errors than lower IQ participants. This may be related to the fact that they emitted more responses in general (although the correlation with response total was not significant). It is also possible that increased behavior within the task brought participants into contact with errors more frequently. Emitting an error and experiencing the associated contingencies shapes behaviors effectively within the task as structured in this investigation. Chain length is also routinely used through the analysis in this paper; the maximum chain length attained by a participant during the first IRA encounter for both chain types was 6-links indicating sufficient range but very few participants exceeded a 2-link response chain. This was a surprise given that Paule (1988b, 1990b, 1999) and Baldwin (2012) reported concerns with a ceiling effect in terms of the length of the chain available to their participants.

Procedural differences unrelated to measurement are also possible sources of divergence between this work and the previous findings. In the current task each IRA session lasted for 5 minutes as opposed to the 15 minute session used by Paule (1988b, 1990b, 1999) and Baldwin (2012). Initially this study attempted to use a 10 minute session for some of the IRA encounters, but it was quickly determined that for the younger children within this age range 10 minutes was too long and the session was shortened to avoid participants abandoning the task. Another difference is the style and extent of instructions given to children participating. In the studies by Paule (1988b, 1990b, 1999) and Baldwin (2012) a video was developed to model the task for the participant, the video include detailed instructions about the various stimuli presented throughout the task. Our participants received extremely minimal instructions prior to the start of the first IRA encounter in order to minimize language comprehension as a source of variance, even during the error correction phase the goal was to prevent repetitive errors, not identify or clarify the rules of the task for the child. In this project an error reset the chain, a reinforcer could never be obtained unless an "errorless response chain" was produced. In Paule and Baldwin's various studies an error did not reset the chain and a reinforcer could be earned when/if the correct response was eventually selected. Differences like this suggest that the procedures selected for the current iteration of the IRA task might be so stringent that they hold even high IQ participants to artificially low levels of performance. Or alternatively, the behavior based advancement criterion develops (or supports) stronger performance in low IQ participants by shaping the responses within the chain.

Another unique contribution made by this study was the inclusion of the DCCS, a card sort task that measures executive function and captures the development in a child's ability to shift through the sets on this task. As expected the DCCS measures showed strong positive correlations with age but did not correlate with IQ, suggesting that these tasks measure independent constructs. The DCCS didn't correlate with Forward and Backward Chaining IRA in the same way, nor did it correlate with Session 1 and Session 3 in the same way. In Session 1 there was a positive correlation between the DCCS and response total for Forward Chaining IRA. In contrast the DCCS correlated with both the response total and PQ for Session 3 Forward Chaining IRA. DCCS didn't correlate with measures of performance for Backward Chaining IRA on either session, only response total and error total. For the DCCS task a high score can be interpreted as an indication of low levels of perseveration. Participants who performed well on the Forward Chaining IRA task also performed well on the DCCS, but the same relationship was not observed for Backward Chaining IRA. This is somewhat surprising as high levels of perseveration (low performance on the DCCS) were expected to be particularly detrimental to performance on the Backward Chaining IRA task and relatively unrelated to the Forward Chaining IRA task. Given the outcome is the reverse of what is expected it's possible that the DCCS might not capture the form of perseveration theorized to be important within the IRA task. For instance the DCCS measures inability to "switch" when verbal instructions are frequent, this might not be the same kind of switching that takes place during the IRA task. The computerized IRA task emphasizes the participants' ability to switch or change a response based on automated feedback after a response rather than verbal instructions prior to a response.

The final element of this study was the regression analysis of the first IRA session, intended to clearly describe the relationship between IQ and the DCCS task for each of the IRA chain types independently. For both chain types the combination of IQ and DCCS accounted for a strikingly small portion of variance. Due to the small sample size, attempts at building a model using multiple variables were omitted.

The results of this study point to the need for additional work to clarify the relationship between the IRA task and IQ. Future research might productively focus on the many procedural differences between this study and early work that identified a strong relationship with IQ by Paule (1988b, 1990b, 1999) and Baldwin (2012). Determining which of these differences, if any, contribute to the different relationship with IQ would help inform the decisions of future investigators. Despite the logistical challenges more information about how IQ relates to an ongoing IRA task would also be interesting. This is particularly true given the widespread use of the IRA task in animal labs as a correlate of human learning and intelligence; in animal labs the procedure is designed and used repeatedly. In addition, comparing the IRA task to alternative measures of executive function would be interesting.

Tables

Table 1

Repeated Measure ANOVA of all relevant IRA factors; for Response Total, Chain Length and PQ.

Effect	Response Total		Chain Length		PQ	
	F	*p*	F	*p*	F	*P*
	Between-Subject					
Order of IRA Experience	0.18	*0.681*	0.42	*0.525*	0.55	*0.468*
Error Correction Group	0.41	*0.532*	1.38	*0.255*	1.35	*0.260*
Order of IRA Experience * Error Correction Group	9.08	***0.007****	4.46	***0.048****	5.32	***0.033****
	Within-Subject					
IRA Task	6.88	***0.017****	2.91	*0.155*	0.87	*0.362*
IRA Task * Order of IRA Experience	0.44	*0.514*	4.74	***0.042****	4.01	*0.060*
IRA Task * Error Correction Group	0.09	*0.763*	0.18	*0.674*	0.70	*0.415*
IRA Task * Order of IRA Experience * Error Correction Group	0.620	*0.441*	0.69	*0.415*	0.21	*0.650*
Session	3.98	***0.012****	4.51	***0.007****	6.85	***0.001****
Session * Order of Experience	3.58	***0.019****	0.75	*0.529*	0.63	*0.598*
Session * Error Correction Group	0.81	*0.491*	2.17	*0.101*	2.21	*0.097*
Session * Order of Experience * Error Correction Group	1.85	*0.149*	1.04	*0.380*	1.15	*0.337*
IRA Task * Session	3.19	***0.030****	2.58	*0.063*	2.49	*0.069*
IRA Task * Session * Order of Experience	4.32	***0.008****	1.47	*0.233*	1.73	*0.171*
IRA Task * Session * Error Correction Group	0.220	*0.882*	0.18	*0.912*	0.37	*0.779*
IRA Task * Session * Order of Experience * Error Correction Group	0.098	*0.961*	0.17	*0.914*	0.08	*0.969*

Notes: Table A11 is an expanded version of this table that shows df, Sum of Squares and Mean Square Error.

* Indicates $p < .05$ ** Indicates $p < .01$

Table 2
RMANOVA of Chain Type when Error Correction Group is removed, sessions 1, 3, 5 and 7 only.

		Df	SS	MS	F	p
Response Total		Between-Subject				
	Order of IRA Experience	1	4690.00	4690.00	2.434	0.132
	Error	23	44318.61	1926.90		
		Within-Subject				
	IRA Task	1	7516.85	7516.85	13.22	0.001*
	IRA Task * Order of IRA Experience	1	1980.27	1980.27	3.48	0.075
	Error	23	13079.50	568.67		
	Session	1	140.81	140.81	0.311	0.582
	Session * Order of IRA Experience	1	2.15	2.15	0.005	0.946
	Error	23	10416.10	452.87		
	IRA Task * Session	1	1022.92	1022.92	6.96	0.015*
	IRA Task * Session * Order of IRA Experience	1	272.80	272.80	1.86	0.186
	Error	23	3381.03	147.00		
Chain Length		Between-Subject				
	Order of IRA Experience	1	1.87	1.87	0.43	0.519
	Error	23	99.97	4.35		
		Within-Subject				
	IRA Task	1	10.13	10.13	5.71	0.025*
	IRA Task * Order of IRA Experience	1	12.13	12.13	6.84	0.015*
	Error	23	40.83	1.78		
	Session	1	9.00	9.00	10.90	0.003*
	Session * Order of IRA Experience	1	0.36	.036	0.44	0.516
	Error	23	19.00	0.83		
	IRA Task * Session	1	5.12	5.21	8.67	0.007*
	IRA Task * Session * Order of IRA Experience	1	0.17	0.17	0.29	0.596
	Error	23	13.83	0.60		
PQ		Between-Subject				
	Order of IRA Experience	1	1.09	1.09	1.20	0.049*
	Error	23	21.05	0.92		
		Within-Subject				
	IRA Task	1	1.77	1.77	5.22	0.032*
	IRA Task * Order of IRA Experience	1	3.01	3.01	8.85	0.007*
	Error	23	7.81	0.34		
	Session	1	3.66	3.66	16.43	0.000*
	Session * Order of IRA Experience	1	0.02	0.02	0.093	0.763
	Error	23	5.12	0.22		
	IRA Task * Session	1	1.02	1.02	4.92	0.037*
	IRA Task * Session * Order of IRA Experience	1	0.02	0.02	0.09	0.763
	Error	23	4.78	0.21		

Notes:
* Indicates $p < .05$
** Indicates $p < .01$

Table 3
Correlations between Forward Chaining IRA Task Session 1 (combined with session 5) and Age, IQ and DCCS

	Age in Months	Full Scale IQ Composite Score	Verbal IQ Composite Score	Performance IQ Composite Score	DCCS Phase	DCCS Correct by Phase Criteria
Response Total	0.279*	0.120	-0.177	0.236	0.362**	0.338*
Error Total	0.011	0.313*	0.101	0.300*	0.111	0.046
Chain Length	0.216	0.003	-0.149	0.089	0.218	0.238
Accuracy	0.252	-0.163	-0.264	-0.025	0.213	0.249
PQ	0.275*	0.060	0.093	0.173	0.214	0.258

Notes: Age n = 52, IQ n = 44, DCCS n = 51
* Indicates $p < .05$
** Indicates $p < .01$

Table 4
Correlations between Backward Chaining IRA Task Session 1 (combined with session 5) and Age, IQ and DCCS

	Age in Months	Full Scale IQ Composite Score	Verbal IQ Composite Score	Performance IQ Composite Score	DCCS Phase	DCCS Correct
Response Total	0.510**	0.269	-0.025	0.365*	0.160	0.122
Error Total	0.239	0.362*	0.112	0.378**	0.104	0.101
Chain Length	0.117	-0.101	-0.077	-0.169	-0.077	-0.068
Accuracy	0.229	-0.092	-0.153	0.012	0.043	0.002
PQ	0.239	-0.041	-0.029	-0.060	-0.034	-0.043

Notes: Age n = 53, IQ n = 46, DCCS n = 53
* Indicates $p < .05$
** Indicates $p < .01$

Figures

Forward Chaining	Backward Chaining
$^{A}1$	$^{C}3$
$^{A}1\,^{B}2$	$^{B}2\,^{C}3$
$^{A}1\,^{B}2\,^{C}3$	$^{A}1\,^{B}2\,^{C}3$

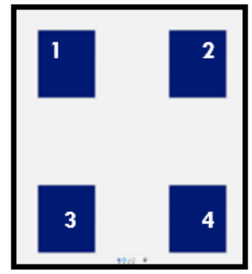
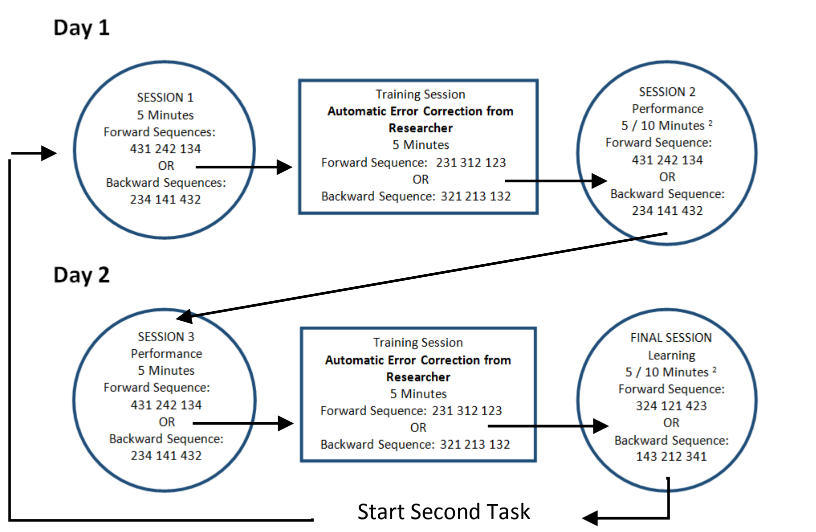
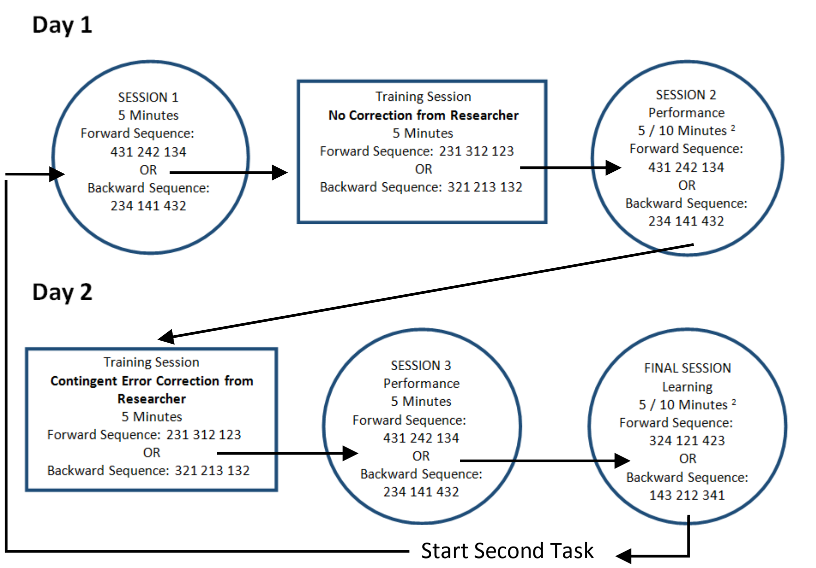


Figure 1. Stimulus Arrangement within a Forward and Backward Chaining IRA Task.
On the right you see a response panel where each response key is labeled. On the left you see a 3 link chain develop from the response keys on the right. Note that the location of the stimuli (A, B and C) is not related to the specific response key required but to the location in the response chain.

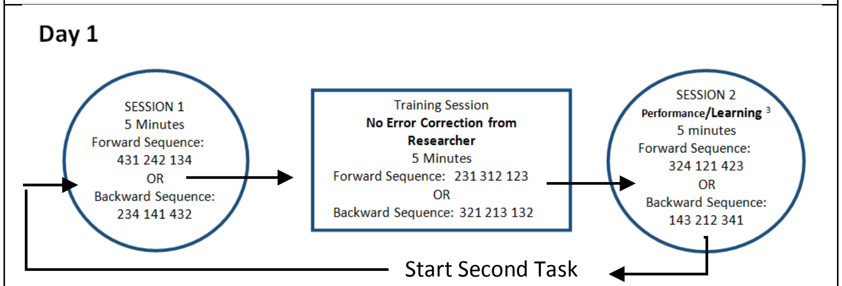
Group 1: Automatic Error Correction n = 15 [1]



Group 2: Contingent Error Correction n = 14 [1]



Group 3: 2 Days Only; No Error Correction n = 25 [1]



Participants experienced the IRA sessions as shown in Fig 2 from left to right/top to bottom. The figure shows days one and two. Then the participant experienced days three and four. Half of the participants began with the backward chain and then received the forward chain. The other half received the reverse sequence. Only data from the sessions represented by circles are displayed in this paper.

[1]For each Group the reported sample size doesn't account for participants missing IQ data, or sessions where data was lost due to a computer glitch. In some analysis the actual n may differ and is reported.

[2]Only data from the first 5 minutes of 10 minute sessions is included in the analysis in this paper.

[3]Due to a program error a few participants in Group 2 and Group 3 experienced a learning session instead of a second or third performance session. Data from these sessions is excluded.

Figure 2. Description of IRA Sessions.

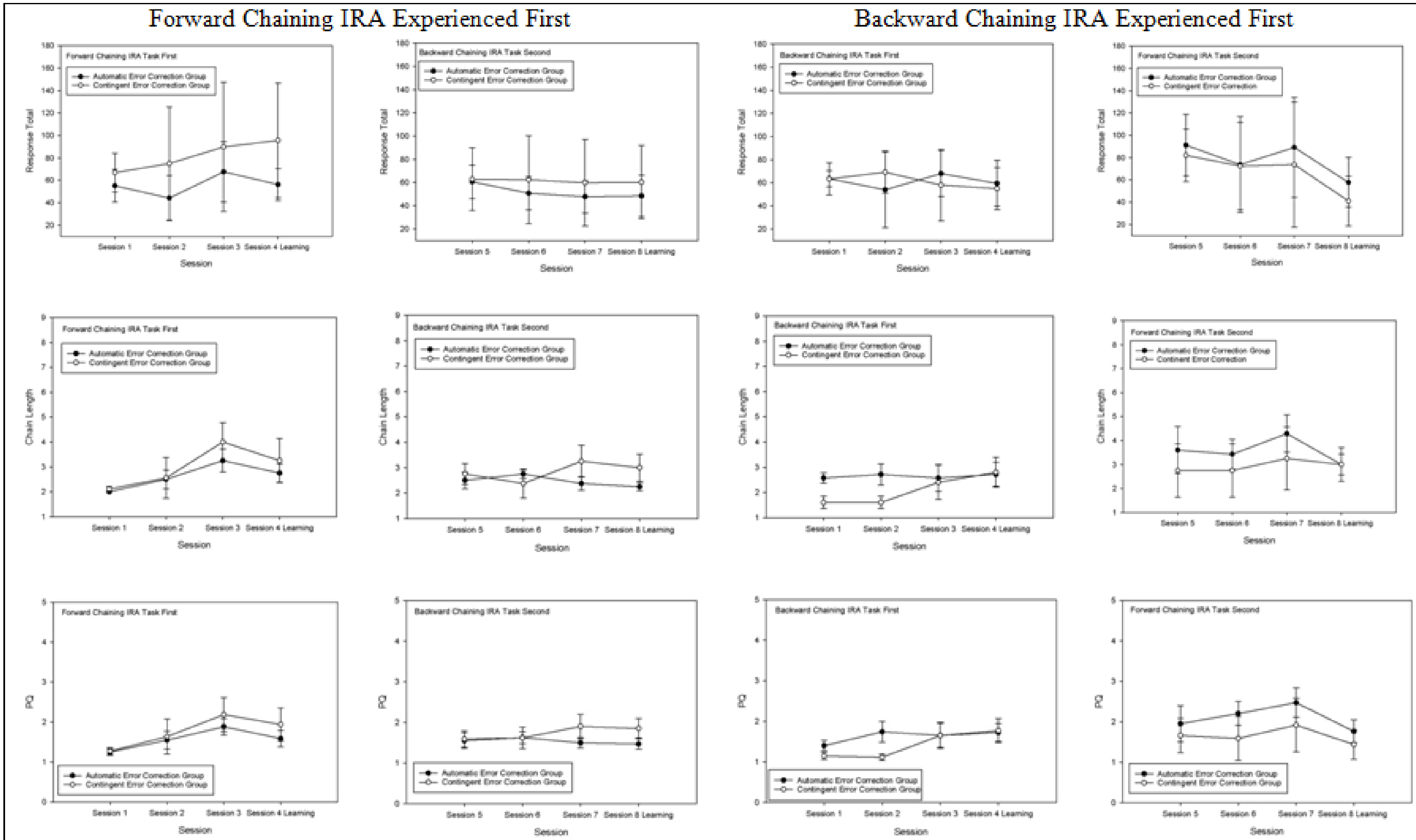


Figure 3. Impact of the Order in which IRA Chain Types are Experienced on Response Total, Chain Length & PQ.
On the left the experience of the experience of the participants who encountered Forward Chaining IRA followed by their experience with Backward Chaining IRA can be read left to right. On the right, again reading left to right, the experience of the participants who experienced. Backward Chaining IRA first, followed by their experience with Forward Chaining IRA.
The four graphs on the top show response total, the 4 graphs in the middle show chain length, and the 4 graphs on the bottom show PQ. The maximum for PQ for this task was 5.0.

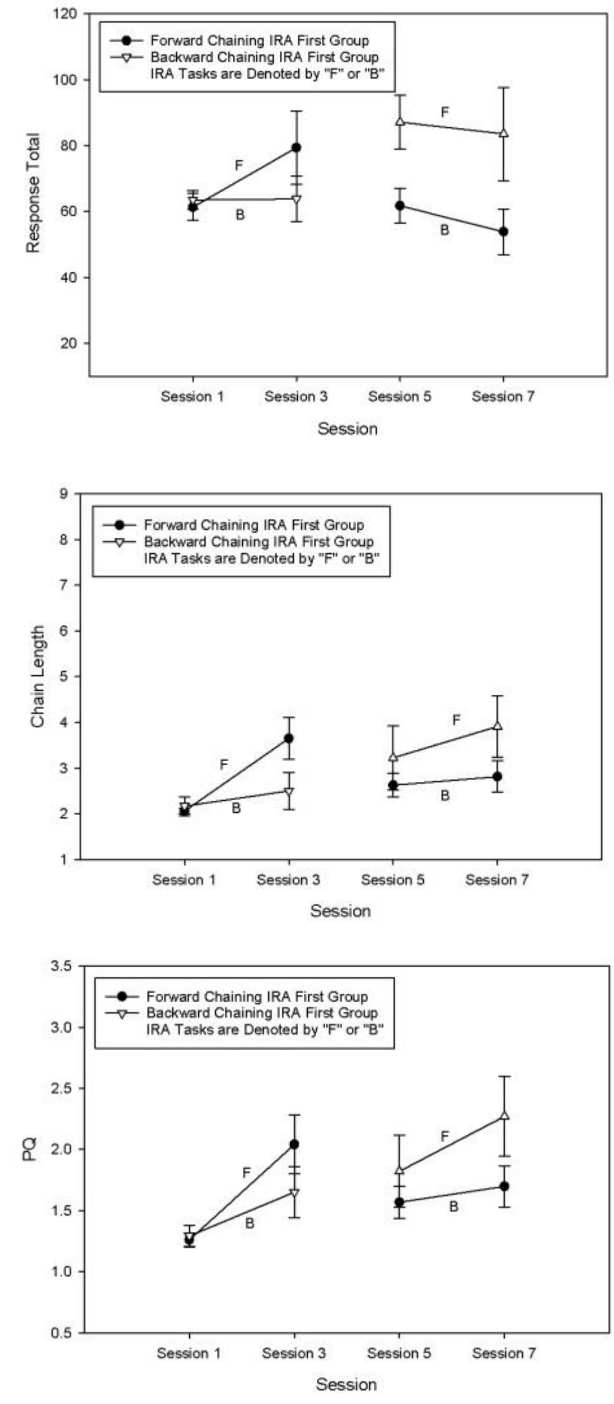


Figure 4

Depiction of change between sessions 1, 3, 5 and 7. The break in the lines indicates that the participants changed IRA tasks, from Forward Chaining to Backward, or the reverse. For clarity each line is marked; those that depict Forward Chaining IRA Tasks have an "F" next to them, lines depicting a Backward Chaining IRA Task have a "B" next to them.

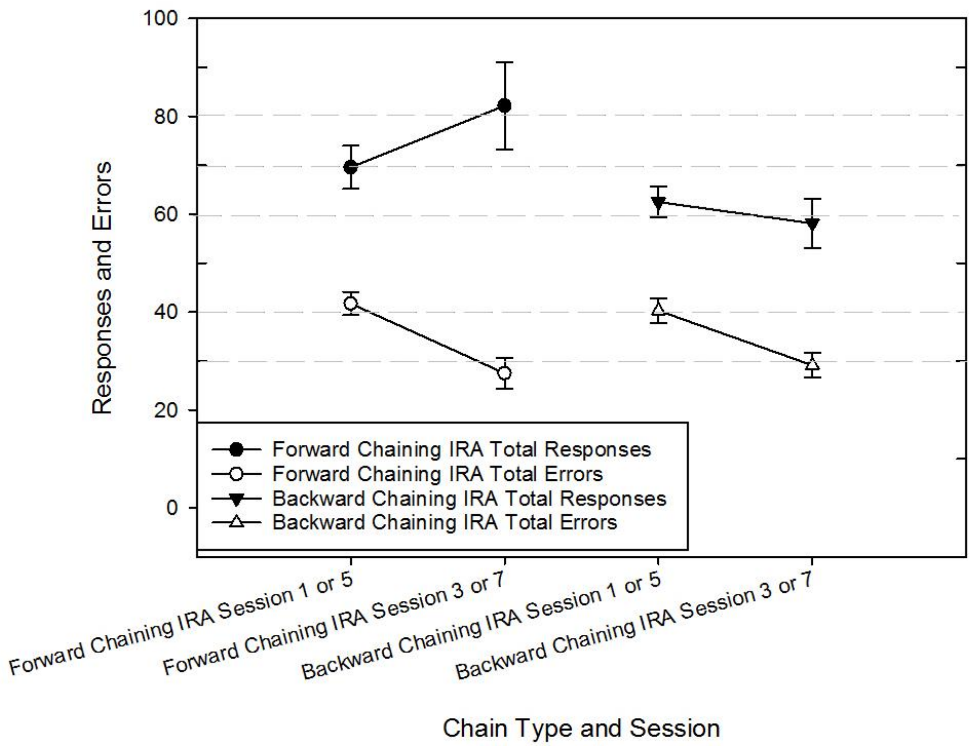


Figure 5 Response Total and Error Total
The response total and error total is shown for the Forward Chaining IRA task on the left, and the Backward Chaining IRA task on the right.

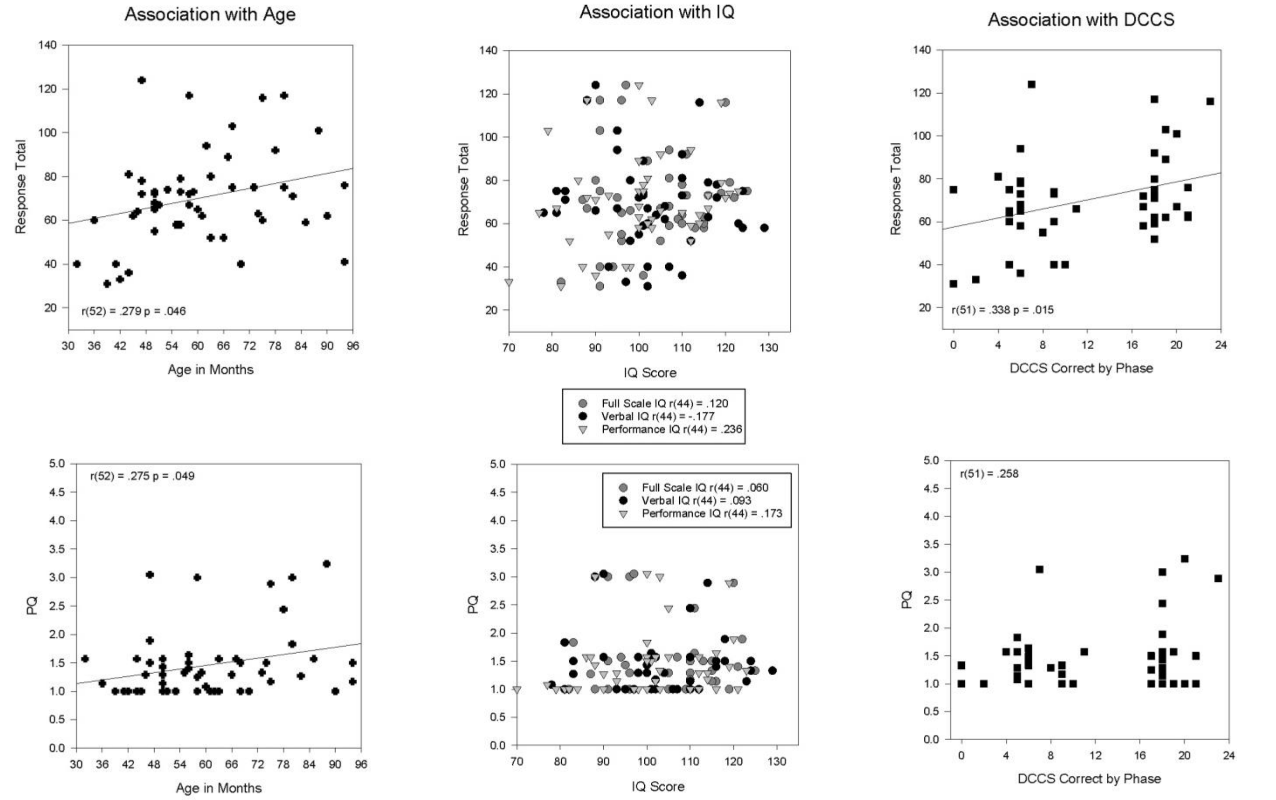


Figure 6. Summary of Associations between Session 1 Forward Chaining IRA and Age, WPPSI—III and DCCS.

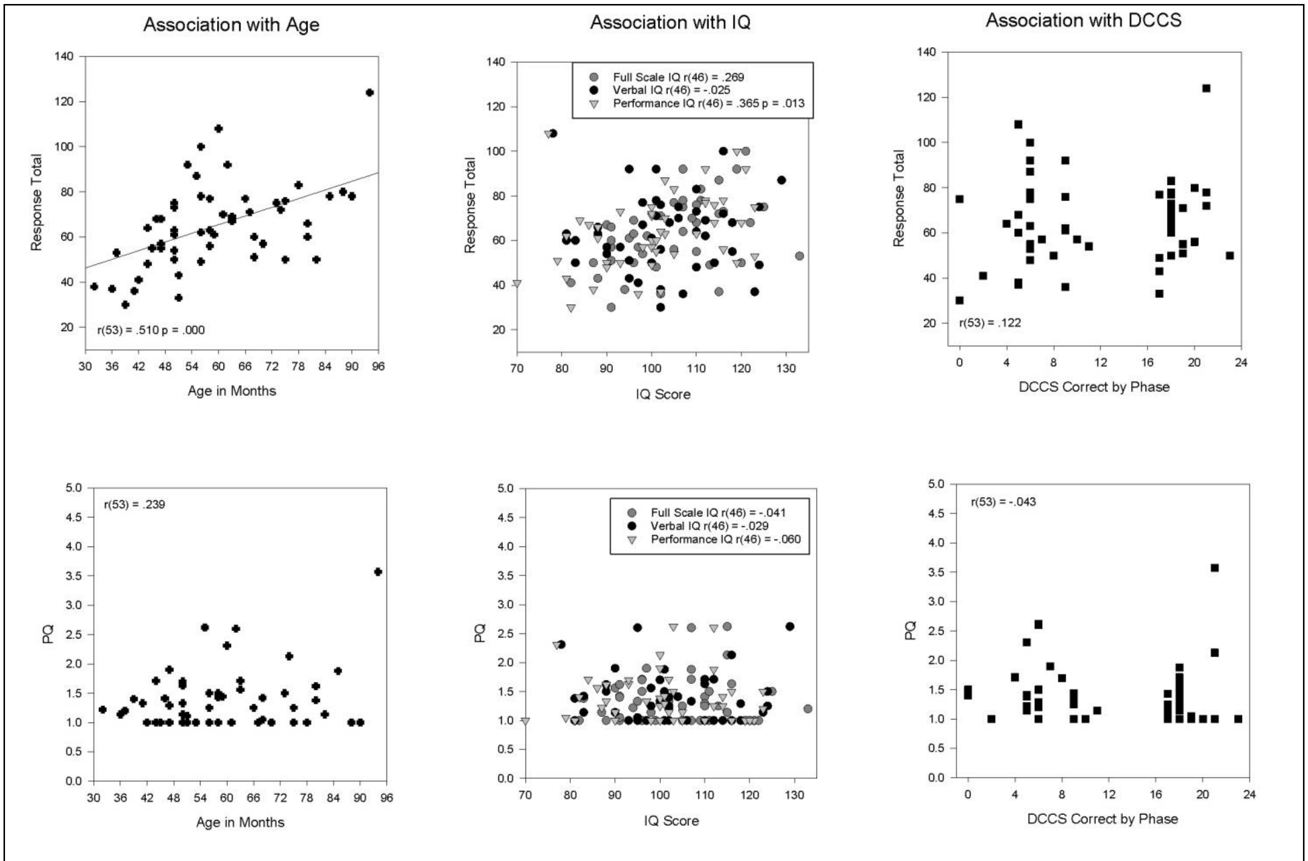


Figure 7. Summary of Associations between First Session Backward Chaining IRA and Age, WPPSI-III and DCCS.

References

Ash , D. W. & Holding, D. H. (1990). Backward versus forward chaining in the acquisition of a keyboard skill. *Human Factors*, *32*(2), 139-146.

Bailey, J.M., Johnson, J.E., & Newland, M.C. (2010). Mechanisms and performance measures in mastery-based incremental repeated acquisition: behavioral and pharmacological analyses. *Psychopharmacology*, 209, 331-341. Doi:10.1007/s00213-010-1801-3

Baldwin, R.L., Chelonis, J.J., Prunty, P.K. & Paule, M.G. (2012). The use of incremental repeated acquisition task to assess learning in children. *Behavioural Processes; 91*, 103-114. DOI:http://dx.doi.org/10.1016/j.beproc.2012.06.004

Batra, M. & Batra, V. (2006). Comparison between forward chaining and backward chaining techniques in children with mental retardation. *The Indian Journal of Occupational Therapy*, XXXVII(3), 57-63.

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*; 99, 15-22.

Boren, J.J. (1963). *The repeated acquisition of new behavioral chains.* Program of the seventy-first annual convent of the American psychological association, Philadelphia Pennsylvania. In the *American Psychologist* 18:7, p 421.

Boren, J.J. & Devine, D.D. (1968). The repeated acquisition of behavioral chains. *The Journal of the Experimental Analysis of Behavior*; 11(6), 651-660.

Catania, A.C. (2007). *Learning; Interim (4th) Edition.* Cornwall-on-Hudson, NY: Sloan Publishing.

Cohn, J., Cox, C. & Cory-Slechta, D.A. (1993). The effects of lead exposure on learning in a multiple repeated acquisition and performance schedule. *Neurotoxicology*; 14, 329-346.

Cox, J.A. & Boren, L.M. (1965). A study of backward chaining. *Journal of Educational Psychology*, 56(5), 270-274.

Deerberry, D. & Rothbart, M.K. (1988). Arousal, Affect and Attention as components of Temperament. *Journal of Personality and Social Psychology*; 55(6), 958-966.

Dinsmore, J.A. (1983). Observing behavior and conditioned reinforcement. *The Behavioral and Brain Sciences*, 6, 693-704.

Fantino, E. (2008) Choice, Conditioned Reinforcement, and the Prius Effect. *The Behavior Analyst*, 31(2), 95-111. doi:

Fantino, E. (1977). Conditioned reinforcement: Choice and information. In W.K. Honig & J.E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 313-339). Englewood Cliffs, NJ: Prentice Hall.

Fantino. E. (1965). Some data on the discriminative stimulus hypothesis of conditioned reinforcement. *The Psychological Record*, 15, 409-415.

Ferster, C.B. & Perrott, M.C. (1968). *Behavior Principles.* United States of America: Meredith Corporation.

Geurts, H.M., Corbett, B. & Solomon, M. (2009) The paradox of cognitive flexibility and autism. *Trends in Cognitive Sciences*; 13(2), 74-82.

Gollub, L.R. (1958). *The chaining of fixed-interval schedules.* (Unpublished doctoral dissertation, Harard University.) As cited in Fantino (2008).

Higgins, S. T. & Woodward, B. M. & Henningfield, J.E. (1989). Effects of atropine on the repeated acquisition and performance of response sequences in humans. *Journal of the Experimental Analysis of Behavior*; 51(1), 5-15.

Hur, J. & Osborne, S. (1993). A comparison of forward and backward chaining methods used in teaching corsage making skills to mentally retarded adults. *The British Journal of Developmental Disabilities*; 39(77), 108-117.
Jurado, M.B. & Rosselli, M. (2007). The exclusive nature of executive functions: A review of our current understanding. *Neuropsychological Review*; 17, 213-233.
Johnson, J. M., Bailey, J. M., Johnson, J.E. & Newland, M.C. (2010). Performance of BALB/c and C57BL/6 mice under an incremental repeated acquisition of behavioral chains procedure. *Behavioral Processes*; 64, 705-714. doi:10.1016/j.beproc2010.04.008
Johnson, K.A. & Senter, R.J. (1965). A comparison of forward and backward chaining techniques for the teaching of verbal sequential tasks. Aerospace Medical Research Lab, Wright-Paterson Air Force Base, Ohio. Retrieved from www.scribd.com on January 6, 2012.
Kaufman, A.S., & Kaufman, N.L. (1990). Kaufman Brief Intelligence test. Circle Pines, MN: American Guidance Services. As cited in Baldwin et al. (2012).
Keehn, J.D. (1967). In bar-holding with negative reinforcement preparatory or perseverative? *Journal of the Experimental Analysis of Behavior*, 10(5), 461-465.
Keller, F.S. & Schoenfeld, W.N. (1950). *Principles of Psychology.* New York: Appleton-Century-Crofts.
Lashley, K.S. (1951). *The problems of serial order in behavior, in Cerebral Mechanisms in Behavior.* Jeffries, L.A. Ed (Wiley, New York).
Lattal, K.A. & Crawford-Goodbey (1985). Homogenous Chains, Heterogeneous Chains and Delay of Reinforcement. *Journal of the Experimental Analysis of Behavior*, 44 (3), 337-342.
Martin & Pear (2007). *Behavior Modification; What it is and how to do it.* Eighth Edition. Upper Saddle River, New Jersey: Pearson, Prentice Hall.
Millenson, J.R. (1967). *Principles of behavioral analysis.* New York: Macmillan.
Neuringer, A., Deiss, C. Olson, G. & College, R. (2000). Reinforced variability and operant learning. *Journal of the Experimental Psychology*; 26(1), 98-111. doi:10.1037//0097-7403.26.1.98
Newland, M.C. & Reile, P.A. (1999). Learning and Behavior Change as Neurotoxic Endpoints. In Tilson, H.A., Harry J. (eds) *Target Organ Series: Neurotoxicology.* Raven Press, New York, pp 311-338.
Neyens, L.G.J. & Aldenkamp, A P. (1996). Stability of Cognitive Measures in Children of Average Ability. *Child Neuropsychology*; 2(3), 161-170. Doi: 0929-7049/97/0303-161$12.00
Paule, M.G. & McMillian, D.E. (1984). Incremental repeated acquisition in the rat: Acute effects of drugs. *Pharmacology Biochemistry & Behavior*; 21, 431-439. doi:0091-3057/84
Paule, M.G., Schulze, G.E. & Slikker, W.J. (1988). Complex brain function in monkeys as a baseline for studying the effects of exogenous compounds. *NeuroToxicology*; 9(3), 463-470.
Paule, M.G., Cranmer, J.M., Wilins, J.D., Stern, H.P. & Hoffman, E.L. (1988b). Quantitation of Complex Brain Function in Children: Preliminary Evaluation Using a Nonhuman Primate Behavioral Test Battery. *NeuroToxicology*; 9(3), 367-378.
Paule, M.G.; Cranmer, J.M. (1990). Complex brain function in children as measured in the NCTR monkey operant test battery. In: Johnson B.L. ed. *Advances in neurobehavioral toxicology: Applications in environmental and occupational health.* Chelsea, MI: Lewis Publishers.
Paule, M.G., Forrester, T.M., Maher, M.A., Cranmer, J.A. & Allen, R.R. (1990b). Monkey Versus Human Performance in the NCTR Operant Test Battery. *Neurotoxicology and Teratology*; 12, 503-507.
Paule, M.G. (1990c). Use of the NCTR Operant Test Batter in Nonhuman Primates. *Neurotoxicology and Teratology*, 12, 413-418.

Paule, M.G., Chelonis, J.J., Buffalo, E.A., Blake, D.J. & Casey, P.H. (1999). Operant Test Battery Performance in Children Correlation with IQ. *Neurotoxicology and Teratology*; 21(3), 223-230.

Pear, J. (2001). *The Science of Learning.* Ann Arbor: Edwards Brothers.

Pearson (2012). *Wechsler Preschool and Primary Scale of Intelligence – Third Edition (WPPSI – III).* Retrieved from http://pearsonassessments.com

Pieper, W.A. (1976). Great apes and rhesus monkeys as subjects for psychopharmacological studies of stimulants and depressants. *Federation Proceedings*; 35(11), 2254-2257.

Pisacreta, R. (1982). A comparison of forward and backward procedures for the acquisition of response chains in pigeons. *Bulletin of the Psychonomic Society*; 20(4), 233-236.

Poling, A., Cleary, J., Berens, K. & Thompson, T. (1990). Neuroleptics and Learning: Effects of Haloperidol, Molindone, Mesoridazine and Thioridazine on the Behavior of Pigeons under a Repeated Acquisition Procedure. *The Journal of Pharmacology and Experimental Therapeutics*; 255(3), 1240-1245. Doi: 0022-3565/90/2553-1240$03.00/0

Rachlin, H. (1935). *Behavior and Learning.* United States of America: W.H. Freeman & Company.

REALbasic 2007 Release 4 [Computer software]. USA: Copyright 1997-2007 REAL Software, Inc.

Schuster, R.H. (1969). A functional analysis of conditioned reinforcement. In D. P. Hendry (Ed.), *Conditioned Reinforcement* (pp. 192-234). Homewood, IL: Dorsey Press.

Shannon, H.E. & Love, P.L. (2004). Within-session repeated acquisition behavior in rats as a potential model of executive function. *European Journal of Pharmacology*, 498, 125-134. doi: 10.1016/j.ejphar.2004.04.054

Sidman, M. (1960). *Tactics of Scientific Research; Evaluating Experimental Data in Psychology.* USA: Authors Cooperative, Inc.

Skinner, B.F. (1938). *The behavior of organisms: An experimental analysis.* New York: Appleton-Century-Crofts.

Slocum, S.K. & Tier, J.H. (2011). An assessment of the efficiency of and child preference for forward and backward chaining. *Journal of Applied Behavior Analysis*; 44(4),793-805.

Smith, G.J. (1999). Teaching a long sequence of behavior using whole task training, forward chaining and backward chaining. *Perceptional and Motor Skills*, 89, 951-965.

Spencer, K., Newland, M.C. (2010a, May). *Low dose Haloperidol Impairs Acquisition of New Sequences During Incremental Repeated Acquisition.* Poster session presented at Association for Behavior Analysis International Annual Convention, San Antonio, TX.

Spencer, K., Newland, M.C. (2010b, May).*Switching incrementing chain type in an Incremental Repeated Acquisition procedure.* Poster session presented at Association for Behavior Analysis International Annual Convention, San Antonio, TX.

Spencer K, Newland, M.C. (2009, April) *Chronic low dose Haloperidol challenges to an incremental repeated acquisition procedure in rats.* Poster session presented at the Auburn University Teaching and Research Festival, Auburn, AL.

Spencer K., Bailey, J. & Newland, M.C. (2008). *Challenges to performance under incremental repeated acquisition schedules: d Amphetamine and transition to a tandem component.* Poster session presented at South Eastern Association for Behavior Analysis, Athens GA.

Spencer Walstrom K, Gillis, J., Newland, M.C. (2011, May) *Designing an Incremental Repeated Acquisition procedure for college students; a series of three studies looking at the relevant parameters.* Paper at Association for Behavior Analysis International Annual Convention, Denver, CO.

Squires, N. (1972). *Preference for conjoint schedules of primary reinforcement and brief-stimulus presentation.* [Unpublished doctoral dissertation, University of California, San Diego.] As cited in Fantino (2008).

Stark, R.E., Tallal, P., Kallman, C. & Mellits, E.D. (1983). Cognitive abilities of language delayed children. *The Journal of Psychology*, 114, 9-19.

Tayler, R.L. (1965). The effect of forward and backward implicit verbal chaining on paired-associate learning. *Psychon*. Sci.; 3, 427-428.

Thompson, D. M. (1970). Repeated acquisition as a behavioral baseline. *Psychon. Sci*. 21(3) 156-157.

Thompson, D. M. (1973). Repeated Acquisition as a Behavioral Baseline for Studying Drug Effects. *The Journal of Pharmacology and Experimental Therapeutics*, 184(2), 506-514.

Thompson, D.M. & Moerschbaecher, J.M. (1979). Drug Effects on Repeated Acquisition. *Advances in Behavioral Pharmacology*, 2, 229-259.

Walls, R.T., Zane, T. & Ellis, W.D. (1981). Forward and Backward Chaining, and Whole Task Methods; Training Assembly Tasks in Vocational Rehabilitation. *Behavior Modification*; 5(1), 61-74.

Weber, N.J. (1978). Chaining Strategies for teaching Sequenced Motor Tasks to Mentally Retarded Adults. *The American Journal of Occupational Therapy*, 32(6), 385-389.

Wechsler, D. (2002a). *WPPSI-III Administration and Scoring Manual.* San Antonio, TX; The Psychological Corporation, A Harcourt Assessment Company.

Wechsler, D. (2002b). *WPPSI-III Technical and Interpretive Manual.* San Antonio, TX; The Psychological Corporation, A Harcourt Assessment Company.

Weinberger, S.B. & Killam, E.K. (1978). Alterations in Learning Performance in the Seizure-Prone Baboon: Effects of Elicited Seizures and Chronic Treatment with Diazepam and Phenobarbital. *Epilepsia*; 19, 301-316.

Weiss, K.M. (1978) A comparison of forward and backward procedures for the acquisition of response chains in humans. *Journal of the Experimental Analysis of Behavior*, 29(2), 255-259.

Wenger, G.R., Schmidt, C., & Davisson, M.T. (2004). Operant conditioning in the Ts65Dn Mouse: Learning. *Behavior Genetics*, 34(1), 105-119. doi: 0001-8244/04/0100-0105/0.

Wright, L.K.M. & Paule, M.G. (2007). Response sequence difficulty in an incremental repeated acquisition (learning) procedure. *Behavioral Processes*; 75, 81-84. doi:10.1016/j.beproc.2007.01.007

Wyckoff, L.B., Jr. (1952). The role of observing responses in discrimination learning: Part I. *Psychological Review*, 59, 431-442.

Zayac, R.M. & Johnston, J.M. (2008). Contriving establishing operations: Responses of individuals with developmental disabilities during a learning task. *Research in Developmental Disabilities;* 29, 202-216.

Zelazo, P.D., Frye, D. & Rapus, T. (1996). An age-related dissociation between knowing rules and using them. *Cognitive Development*, 11, 37-63.

Zelazo, P.D. (2006). The dimensional Change Card Sort (DCCS): a method of assessing executive function in children. *Nature Protocols*, 1(1), 297-302. Doi:10.1038/nprot.2006.46

Appendix A
Tables

Table A1
Evolution of the repeated acquisition procedure.

Study	Procedure Type	Major Contributions	Unique Methods of Measurement	Error total	Accuracy	Chain Length Reached
Boren and Devine, 1968	RA	Initial description of the repeated acquisition procedure	Performance errors Learning errors	X		
Thompson, 1970	RA	Simplified shaping procedure		X	X	
Pieper, 1976	IRA*	Incrementing chain with a fading prompt (stimulus light).				X
Weinberger and Killam, 1978	IRA	Incrementing chain with advancement criteria.	Time to criterion function Efficiency	X		
Thompson and Moerschbaucher, 1979	IRA	Defining the performance and learning phases.	Responses/minute, Percent errors, Correct responses on a cumulative record	X		

Notes: *Pieper 1976 might not be considered the first IRA procedure because a tandem schedule was the final result of the fading procedure.

Table A2
Procedural variations in the animal lab.

Study	Procedure & Chain Type	Species	Stimuli	Chain Length	# of Levers	Criterion for Chain Advancement	Schedule of Reinforcement	Error Procedure	Session Length	Maximum # reinforcers
Boren and Devine, 1968	RA – Whole Chain	Rhesus monkeys	Lights above groups of Levers	4	12	N/A	FR 5 for each response element within the chain	Time out lengthened by additional responses; did not reset the chain	Undefined, ended when animal attained 70 reinforcers	70
Thompson, 1970	RA – Whole Chain	Pigeons	Colored Key Lights	4	4	N/A	FR 1 sequence	Time out, did not reset the chain	N/A	60
Pieper, 1976	IRA* - Backward Chaining with Shaping	Great apes and rhesus monkeys	Cue lights faded to a Tandem Schedule	6	6	Successive correct responses needed to fade out the cue lights through 8 levels of brightness. This fading procedure could increase with errors as well.	FR 1 sequence	Time out; unspecified if this reset the chain or not.	50 minutes or Successful 6 lever sequence	No Maximum
Weinberger and Killam, 1978	IRA – Backward Chaining	Baboons	Colored cue lights.	5	3	Starting at a 2 link chain, 21 consecutive errorless chains required for advancement.	FR 1 Sequence	Error light, no response time out, chain not reset	Failure to reach 21 errorless chains in 15 minutes or no responses for 2 minutes.	No Maximum
Thompson and Moerschbaucher, 1979	IRA – Backward Chaining	Monkeys	Geometric forms projected on press plates.	4	3	Not Described	Various Schedules used – FR 5 to FR 50 and a VR 3 Tested	Errors produced a time out but did not reset the chain.	Multiple Schedule – changed after 15 minutes or 10 reinforcers - 30 minute MAX	20
Paule & McMillian (1984)	IRA – Backward Chaining	Rats	Lights unrelated to lever position	5	3	40 correct sequences	FR 1 Sequence	1 second illum. of incorrect light and son-alert, did not reset the chain.	1 hour or 40 correct responses on the 5 link chain	240

Table A2 Continued

Study	Procedure & Chain Type	Species	Stimuli	Chain Length	# of Levers	Criterion for Chain Advancement	Schedule of Reinforcement	Error Procedure	Session Length	Maximum # reinforcers
Poling, Cleary, Berens & Thompson 1990	Repeated Acq. – Backward Chaining across training sessions.	Pigeons	Color of key light.	4	3	N/A	FR 5 Sequence	Unspecifie d.	1 hour or 50 chains	Not Described
Cohn, Cox & Cory-Slechta, 1993	Repeated Acquisition	Rats	Not Identified	3	3	N/A	FR1 – FR2 Sequence	Errors increased the reinforcemen t schedule to an FR2	1 hour	100
Wenger, Schmidt & Davvison, 2004	IRA – Backward Chaining	Mice	Not Identified	4	3	5 correct responses, 10 correct responses, 30 correct responses, 20 correct responses	FR 1Sequence	5 second time out and reset chain	65 sr+ or 3600 seconds	65
Spencer and Newland, 2010a	IRA – Forward Chaining & Backward Chaining	Rats	Tones	4	3	10 consecutive correct responses; 5 consecutive correct responses; 5 consecutive correct responses	FR 1 Sequence	Time out, reset the chain	60 min or 50 reinforcers in the 4 link chain	No Max
Bailey, Johnson and Newland, 2010	IRA – Forward Chaining and Backward Chaining	Rats	Tones	4	3	10 consecutive correct responses; 5 consecutive correct responses; 5 consecutive correct responses	FR 1 Sequence	Time out, reset the chain	60 min or 50 reinforcers in the 4 link chain	No Max
Johnson et al. 2010	IRA – Forward Chaining & Backward Chaining	Mice	Tones	6	3	10 consecutive correct responses; 5 consecutive correct responses; 5 consecutive correct responses	FR 1 Sequence	Time out, reset the chain	60 min or 50 reinforcers in the 4 link chain	No Max

Notes: *Pieper 1976 might not be considered the first IRA procedure because a tandem schedule was the final result of the fading procedure.

Table A3
IRA component variables in studies with human subjects.

Study	Type of Chain	Length of Chain	# of Levers	Criterion for Chain Advancement	Stimuli to establish chain schedule	Component and Task Exposure	Length of Session	Maximum # Reinforcers	Age Range Tested	N
Paule, Cranmer, Wilkins, Stern & Hoffman (1988b)	Backward	6	4	3 correct responses	Multiple Lights	Single Measurement ?	15 minutes	18 – Nickels & Correct Light	3-11 years	20
Paule & Cranmer (1990)	Presents additional results from Paule, Cranmer, Wilkins, Stern & Hoffman (1988b).									
Paule, Forrester, Maher, Cranmer & Allen, 1990b	Backward	6	4	3 correct responses	Multiple Lights	Single Measurement	15 minutes	18 – Nickels & Correct Light	4-8 years	71
Paule et al. (1999)	Backward	6	4	3 correct responses	Multiple Lights	Single Measurement	15 minutes	18 – Nickels & Correct Light	6 years	115
Zayac & Johnston, 2008	Unclear	4	4	3 correct **consecutive** responses	Background screen color	Learning Component	Unlimited	No Maximum – Access to a video game	37-43 years	3
Spencer Walstrom, Gillis and Newland, 2011	Forward & Backward chaining groups	9	4	3 correct **consecutive** responses	Tones	Single Measurement, Learning & Performance Components Tested	Two 25 minute sessions or One 40 minute session	No Maximum – Screen Saying Good Job	18-25 years	39 & 58
Baldwin, Chelonis, Prunty & Paule (2012)	Backward Chaining	6	4	3 correct responses (No consecutive requirement)	Colored lights	Single Measurement	15 minutes	18 - Nickels	5-13 years	837

Notes:

Table A4
Literature comparing forward and backward chaining methods.

Study	Task Taught via Chaining	Participants	Superior Method	Outcome is based on	Notes
Outcome based on performance during acquisition.					
Johnson & Senter, 1965	Verbal List Learning (Nouns or Numbers)	College Students (n= 24)	1 No Difference 2 Forward chaining 3 Forward Chaining	1 Length of chain learned and recall accuracy Immediately after criterion , 2 # of trials to reproduce the list 3 Number of Errors	Three experiments were described. Retrieved from scribd.com – several portions are redacted.
Wilcox (1974)	Paper Folding & Numerical Procedures	Female College Students (n=176)	No Difference	Acquisition and Retention (one week delay)	Examined each task at a short, medium and long chain length. Backward chaining was superior in one of the six comparisons (long numerical procedures).
Weber (1978)	Assembly of plastic pieces from the Remco Science Kit (Mechanical Pysicis, Style 416)	Mentally Retarded Adults with age range from 22 to 60(n=24)	Backward Chaining superior in terms of learning time. No Difference in terms of retention performance.	Learning time (number of errors prior to reaching criterion) and Retention Performance (delayed by 20 hours)	
Weiss (1978)	6-link sequence on 4 computerized buttons.	College Students (n=11)	Forward Chaining	True Errors (Only those occurring after an initial correct response on each link of the chain.)	
Walls, Zane & Ellis (1981)	Assembly of a bicycle brake, a meat grinder, and a carburetor.	Vocational rehabilitation clients (mild to moderate MR) (n=22)	No Difference	Number of Errors	
Pisacreta (1982)	4-link response chains on a nine-key panel.	Pigeons (n=4)	No Difference *Backward chaining first group outperformed the Forward chaining first group on both types of chaining.*	Number of errors. Number of sessions to acquire the chain.	Within subject comparison, all animals made fewer errors in whichever procedure they experienced first.
Spooner & Spooner (1984)			Backward Chaining		Exercise Literature Currently a Secondary Source – Article Ordered

Table A4 Continued

Study	Task Taught via Chaining	Participants	Superior Method	Outcome is based on	Notes
Ash & Holding (1990)	Musical Keyboard Skills	College Students (n=61)	Forward Chaining	Number and type of error.	Forward chaining was superior to backward chaining. Backward chaining was superior to whole task presentation in terms of accuracy. Forward chaining was superior to whole task presentation in terms of accuracy and timing consistency.
Hur & Osborne (1993)	Corsage making skills.	Children and adults with MR age range from 19 to 47 (n=20).	No Difference	Accuracy during training, accuracy at a follow-up session delayed by one week.	The follow up session occurred after the participant attained mastery at 100% accuracy.
Smith (1999)	120 step physical movement	College Students (n=75)	Forward Chaining	Number of Errors	Included a comparison of whole task training (always inferior).
Batra & Batra (2006)	Putting on Socks and Shoes	Children with MR (n=42)	No Difference	Number of errors, prompts and reinforcers	
Outcome based on speed of training.					
Cox & Boren, 1965	Military Procedural Task (Missile preparation)	Military Personnel (n=30)	No Difference	Duration of training to proficiency (100% accuracy on the task 1 time).	
Johnson & Senter, 1965	Verbal List Learning (Nouns or Numbers)	College Students (n= 24)	1 No Difference 2 Forward chaining 3 Forward Chaining	1 Length of chain learned and recall accuracy Immediately after criterion , 2 # of trials to reproduce the list 3 Number of Errors	Three experiments were described. Retrieved from scribd.com – several portions are redacted.
Pisacreta (1982)	4-link response chains on a nine-key panel.	Pigeons (n=4)	No Difference *Backward Chaining first group outperformed the Forward chaining first group on both types of chaining.*	Number of errors. Number of sessions to acquire the chain.	Within subject comparison, all animals made fewer errors in whichever procedure they experienced first.

Table A4 Continued

Study	Task Taught via Chaining	Participants	Superior Method	Outcome is based on	Notes
Slocum & Tiger (2011)	Motor sequence 3, 6, 9 or 18 steps long.	Children with DD age 11 or 12 years (n=4)	No Difference	Trials to mastery,	Three studies; differential sensitivity to simple and complex tasks, and child preference via concurrent-chains preference assessment.
Outcome based on retention of the skill.					
Johnson & Senter, 1965	Verbal List Learning (Nouns or Numbers)	College Students (n= 24)	1 No Difference 2 Forward chaining 3 Forward Chaining	1 Length of chain learned and recall accuracy Immediately after criterion , 2 # of trials to reproduce the list 3 Number of Errors	Three experiments were described. Retrieved from scribd.com – several portions are redacted.
Wilcox (1974)	Paper Folding & Numerical Procedures	Female College Students (n=176)	No Difference	Acquisition and Retention (one week delay)	Examined each task at a short, medium and long chain length. Backward chaining was superior in one of the six comparisons (long numerical procedures).
Weber (1978)	Assembly of plastic pieces from the Remco Science Kit (Mechanical Pysicis, Style 416)	Mentally Retarded Adults with age range from 22 to 60(n=24)	Backward Chaining superior in terms of learning time. No Difference in terms of retention performance.	Learning time (number of errors prior to reaching criterion) and Retention Performance (delayed by 20 hours)	
Hur & Osborne (1993)	Corsage making skills.	Children and adults with MR age range from 19 to 47 (n=20).	No Difference	Accuracy during training, accuracy at a follow-up session delayed by one week.	The follow up session occurred after the participant attained mastery at 100% accuracy.

Table A4 Continued.

Study	Task Taught via Chaining	Participants	Superior Method	Outcome is based on	Notes
INDIRECT COMPARISON					
Tayler (1965)	Paired-associative chains (word learning).	College students (n=120)	Forward Chaining	# of correct anticipations of the implicit portion of the chain.	This study is placed under indirect comparisons because the task, implicit associations may not meet technical definitions of forward and backward chaining, as not all portions of the chain are explicitly experienced during training.
Keehn (1967)	Shock avoidance responding	Rats (n=?)	Forward Chaining	Number of subjects successfully emitting the chain after training.	Explicitly states no attempt is made to establish the relative merits of procedure types.

Notes: A direct comparison is one in which the primary research question compared forward and backward chaining. * No Difference indicates the lack of a statistically significant OR a visually meaningful difference, not that there was in fact 0 differences between outcome variables. In most studies extremely small differences existed which were not consistently in favor of one form of chaining or the other. ** Population descriptions are directly from the article as opposed to using language appropriate today.

Table A5
Dimensional Change Card Sort (DCCS) task details.

Trial and Card	Left Target Blue Rabbit	Right Target Red Boat	Phase	Instructions	Feedback
				"Here's a blue rabbit and here's a red boat. Now, we're going to play a card game. This is the color game. In the color game, all the blue ones go here [point] and all the red ones go there [point]."	
Example 1 – Blue Boat	Correct			"See, here's a blue one. It goes here." [Place Boat]. "If it's blue it goes here [point], but if it's red it goes there [point]."	N/A
Example 2 – Red Rabbit		Correct		"Now here's a red one. Where does it go?"	Praise for correct response OR Correct error.
1 – Blue Boat	Correct		Phase 1	Repeat Rule.	Always be neutral – no praise.
2 – Red Rabbit		Correct	Phase 1	Repeat Rule.	Always be neutral – no praise.
3 – Blue Boat	Correct		Phase 1	Repeat Rule.	Always be neutral – no praise.
4 – Red Rabbit		Correct	Phase 1	Repeat Rule.	Always be neutral – no praise.
5 – Blue Boat	Correct		Phase 1	Repeat Rule.	Always be neutral – no praise.
6 – Red Rabbit		Correct	Phase 1	Repeat Rule.	Always be neutral – no praise.
				(Repeat 2 TIMES) "Now we're going to play a new game. We're not going to play the color game. We're going to play the shape game. In the shape game all the rabbits go here [pointing], and all the boats go there [pointing]. Remember, if it's a rabbit, put it here, but if it's a boat put it there. OK?"	
7 – Blue Boat		Correct	Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
8 – Red Rabbit	Correct		Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
9 – Blue Boat		Correct	Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
10 – Red Rabbit	Correct		Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
11 – Blue Boat		Correct	Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
12 – Red Rabbit	Correct		Phase 2	"Now, this is a (label shape) where should it go.	Always be neutral – no praise.
				"OK, you played really well. Now I have a more difficult game for you to play. In this game, you sometimes get cards that have a black border around it like this one [show red rabbit with a border]. If you see cards with a black border, you have to play the color game. In the color game, red ones go here [point] and blue ones go there [point]."	
Example 3B- Red Rabbit with Border		Correct		"This card's red, so I'm going to put it right there [place card]."	Always be neutral – no praise.
Example 4 – Red Rabbit	Correct			"But if the cards have no black border, like this one [show red rabbit without a border] you have to play the shape game. In the shape game, if it's rabbit, we put it here, but if it's a boat, we put it there [pointing to tray]. This one's a rabbit, so I'm going to put it right here [place card in tray]. OK? Now it's your turn."	Always be neutral – no praise.
13B – Blue Boat	Correct		Phase 3	Repeat the rules, "If there's a border, play the color game. If there's no border, play the shape game." LABEL THE CARD AS having a border or not.	Always be neutral – no praise.
14 – Red Rabbit	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
15B – Blue Boat	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
16 – Blue Boat		Correct	Phase 3	Same as trial 13.	Always be neutral – no praise.
17B – Blue Boat	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
18 – Red Rabbit	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
19B – Red Rabbit		Correct	Phase 3	Same as trial 13.	Always be neutral – no praise.
20 – Blue Boat		Correct	Phase 3	Same as trial 13.	Always be neutral – no praise.
21B – Blue Boat	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
22 – Red Rabbit	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.
23B – Red Rabbit		Correct	Phase 3	Same as trial 13.	Always be neutral – no praise.
24 – Blue Boat	Correct		Phase 3	Same as trial 13.	Always be neutral – no praise.

Note: Cards with a border are denoted with a B after the trial #.

Table A6
Description of Data Collection Locations

Research Site	Location	Date Range	# of Participants who Completed the Study	# of Potential Participants who couldn't be accommodated
Location A	A Day Care in **Opelika Alabama**	April 2010 to May 2010	6	10
Location B	A Day Care in **Anchorage Alaska**	August 2011 to November 2011	9	0
		January 2013 to August 2013	8	0
Location C	A Karate Studio in **Fresno California**	November 2014	13	0
Location D	Public Locations in **Fresno California**	November 2014	6	0
Location E	A Day Care in **Madera California**	November 2014	8	0
Location F	An ABA Clinic in **Lake Charles Louisiana**	December 2014 to January 2015	4	0

Notes: In order to protect the confidentiality of the participants the actual names of research sights are not provided.

Table A7
Socio demographic information.

Family Annual Income		Race		Daily TV Viewing		Daily Video Game Exposure	
<$25,000	7	Caucasian	35	< 1 hour of TV	15	No Time Playing Video Games	21
$25,000 - $50,000	13	African American or Biracial	5	1-3 hours of TV	34	<1 hour	28
$50,000 - $75,000	10	Hispanic	10	4-10 hours of TV	6	1-8 hours	6
$75,000- $100,000	3	Asian	0				
$100,000- $125,000	11	Other – Fill In	2				
$125,000-$150,000	5	Left Blank	3				
>$150,000	4						
Left Blank	2						

Languages Spoken*		Learning Disabilities and Special Services*			
More than one language frequently spoken in their home	3 (Spanish)	ADHD	2	Applied Behavior Analysis	3
English is the primary language in the home with other languages spoken occasionally.	6 (Spanish), 1 (Arabic), 1 (Greek)	Autism Spectrum Disorder	3	Occupational Therapy	2
Child is fluent in language other than English	5 (Spanish)	Speech Therapy	6	Some medication reported	8*
		Physical Therapy	2		

Notes: N=55, *The categories listed under Language Exposure and Learning Disabilities and Special Services are not exclusive; for instance some of the participants with more than one language spoken in the home were also reported as being fluent in another language. When Language Spoken was examined in relation to study variables the participants a single variable called "Language Exposure" was formed to combine participants with various exposure to Spanish (n=9) and exposure to other languages (n=2). The same is true of Learning Disabilities and Special Services, participants who reported a diagnosis or special service were combined (n=8).

Medications reported: Three of these participants reported using Claritan, and two reported using Flovent, other medications reported by at least one participant included Advair, Singular, Albuterol, a generic "inhaler", Nasonex, Miralax, Concerta and Strattera.

Table A8
Description of Error Correction Procedure

Location in Chain	Condition	Investigator Response
1 Link Chain	After 1 incorrect response.	“Try a different button.”
	After correct responses.	*Praise and label the behavior that was correct.* “You pushed the (button) and found (character).”
	After repeating the same incorrect response more than once. OR After multiple incorrect responses on all response keys. OR When the participant stopped responding. OR when the participant asked for help.	Gestured / model or physically prompt to a different button. *The investigator **never** told the participant the correct response “You have to push the (button).”*
2 Link Chain or Longer	After 1 incorrect response	“Try a different button.”
	After correct responses	*Praise and label the behavior that was correct.* “You pushed the (button) and found (character).”
	After repeating the same incorrect response more than once. OR After multiple incorrect responses on all response keys. OR When the participant stopped responding. OR when the participant asked for help.	“I hear a new/old sound. That means we push ….” Gestured / model or physically prompt to a different /or **correct** button. *The investigator **only** indicated a specific correct response **if** it was one learned in a previous iteration of the task.*

Notes:

Table A9

Description of IRA Performance Across Sessions when sample is divided by Error Correction Group and Order of IRA Experience.

		Session 1			Session 2 Performance Sequence			Session 3 Performance Sequence			Final Learning Session		
		n	Mean	SD	N	Mean	SD	N	Mean	SD	n	Mean	SD
Forward Chaining IRA Task													
Response Total	Forward First Group 1	8	**55.13**	14.60	8	**44.13**	20.09	8	**67.63**	26.90	8	**56.27**	14.16
	Forward First Group 2	9	**66.89**	17.37	9	**74.93**	50.47	9	**89.81**	57.48	8	**95.63**	51.13
	Forward Second Group 1	5	**91.20**	27.62	7	**73.71**	43.61	7	**89.14**	44.85	7	**57.74**	22.35
	Forward Second Group 2	4	**82.00**	23.61	4	**72.50**	39.24	4	**73.75**	55.96	5	**50.20**	28.24
Error Total	Forward First Group 1	8	**36.5**	10.14	8	**23.00**	7.80	8	**28.50**	5.61	8	**29.07**	8.69
	Forward First Group 2	9	**45.44**	14.64	9	**37.93**	23.60	9	**30.59**	24.62	8	**40.88**	28.00
	Forward Second Group 1	5	**41.80**	12.54	7	**26.29**	15.52	7	**24.29**	13.10	7	**28.56**	12.08
	Forward Second Group 2	4	**45.00**	12.30	4	**33.50**	17.82	4	**26.50**	14.20	5	**20.00**	15.41
Chain Length	Forward First Group 1	8	**2.00**	0.00	8	**2.50**	1.07	8	**3.25**	1.28	8	**2.75**	1.04
	Forward First Group 2	9	**2.11**	0.33	9	**2.56**	2.46	9	**4.00**	2.35	8	**3.25**	2.49
	Forward Second Group 1	5	**3.60**	2.19	7	**3.43**	1.62	7	**4.29**	2.06	7	**3.00**	1.15
	Forward Second Group 2	4	**2.75**	2.22	4	**2.75**	2.12	4	**3.25**	2.63	5	**3.60**	1.82
Accuracy	Forward First Group 1	8	**0.34**	0.04	8	**0.44**	0.15	8	**0.52**	0.18	8	**0.47**	0.11
	Forward First Group 2	9	**0.34**	0.67	9	**0.42**	0.23	9	**0.57**	0.24	8	**0.54**	0.25
	Forward Second Group 1	5	**0.48**	0.26	7	**0.59**	0.23	7	**0.65**	0.24	7	**0.47**	0.20
	Forward Second Group 2	4	**0.40**	0.24	4	**0.44**	0.29	4	**0.50**	0.29	5	**0.55**	0.27
PQ	Forward First Group 1	8	**1.25**	0.23	8	**1.55**	0.63	8	**1.88**	0.58	8	**1.59**	0.57
	Forward First Group 2	9	**1.27**	0.24	9	**1.63**	1.32	9	**2.19**	0.58	8	**1.94**	1.17
	Forward Second Group 1	5	**1.95**	1.00	7	**2.20**	0.78	7	**2.47**	0.96	7	**1.76**	0.75
	Forward Second Group 2	4	**1.66**	0.85	5	**1.59**	1.09	4	**1.92**	1.33	5	**1.76**	0.95

Table A9 Continued
Backward Chaining IRA

		Session 1			Session 2 Performance Sequence			Session 3 Performance Sequence			Final Learning Session		
		n	Mean	SD	n	Mean	SD	N	Mean	SD	n	Mean	SD
Response Total	Backward First Group 1	8	**60.63**	14.53	8	**50.75**	14.24	8	**47.75**	14.09	8	**48.40**	17.88
	Backward First Group 2	8	**62.88**	27.22	8	**62.38**	37.98	8	**59.88**	37.26	8	**38.85**	13.86
	Backward Second Group 1	7	**63.57**	7.07	7	**54.0**	32.79	7	**68.00**	19.78	7	**59.50**	19.96
	Backward Second Group 2	5	**63.40**	13.97	5	**69.20**	18.27	5	**58.00**	31.07	4	**41.00**	22.35
Error Total	Backward First Group 1	8	**37.75**	6.01	8	**28.69**	8.86	8	**28.75**	9.13	8	**29.17**	8.31
	Backward First Group 2	8	**34.75**	18.55	8	**31.00**	20.21	8	**20.38**	9.38	8	**31.00**	13.86
	Backward Second Group 1	7	**44.00**	12.14	7	**28.86**	15.59	7	**37.14**	14.75	7	**34.57**	12.22
	Backward Second Group 2	5	**48.00**	12.14	5	**54.20**	16.10	5	**32.40**	17.90	4	**21.50**	17.37
Chain Length	Backward First Group 1	8	**2.50**	0.93	8	**2.75**	0.46	8	**2.38**	0.74	8	**2.25**	0.46
	Backward First Group 2	8	**2.75**	1.16	8	**2.38**	1.60	8	**3.25**	1.75	8	**3.50**	1.93
	Backward Second Group 1	7	**2.57**	0.53	7	**2.71**	1.11	7	**2.57**	1.40	7	**2.71**	1.25
	Backward Second Group 2	5	**1.60**	0.55	5	**1.60**	0.55	5	**2.40**	1.52	4	**3.00**	1.41
Accuracy	Backward First Group 1	8	**0.35**	0.15	8	**0.43**	0.10	8	**0.39**	0.15	8	**0.37**	0.10
	Backward First Group 2	8	**0.43**	0.15	8	**0.43**	0.17	8	**0.58**	0.18	8	**0.47**	0.21
	Backward Second Group 1	7	**0.31**	0.15	7	**0.41**	0.17	7	**0.42**	0.24	7	**0.40**	0.14
	Backward Second Group 2	5	**0.25**	0.06	5	**0.22**	0.05	5	**0.41**	0.23	4	**0.48**	0.25
PQ	Backward First Group 1	8	**1.55**	0.52	8	**1.62**	0.41	8	**1.50**	0.36	8	**1.47**	0.36
	Backward First Group 2	8	**1.59**	0.56	8	**1.62**	0.75	8	**1.89**	0.86	8	**2.06**	0.82
	Backward Second Group 1	7	**1.40**	0.34	7	**1.74**	0.68	7	**1.65**	0.78	7	**1.72**	0.58
	Backward Second Group 2	5	**1.14**	0.20	5	**1.11**	0.18	5	**1.65**	0.72		**1.44**	0.72

Note:

Table A10
Description of Participants within Each Cell

	N	Age (Months)	Full Scale IQ
Forward Chaining First –Group 1 Automatic Error Correction	8	53.25	101.25
Forward Chaining First – Group 2 Contingent Error Correction	9	61.78	109.33 (n = 6)
Backward Chaining First – Group 1 Automatic Error Correction	7	63.00	105.71
Backward Chaining First – Group 2 Contingent Error Correction	5	62.60	115 (n = 2)

Note:

Table A11
Comprehensive Repeated Measure ANOVA.

IRA Variable	Source	df	SS	MS	F	p
				Between-Subject		
Response Total	Order of IRA Experience	1	345.04	345.04	0.18	0.681
	Error Correction Group	1	800.32	800.322	0.41	0.532
	Order of IRA Experience * Error Correction Group	1	17920.07	17920.07	9.08	**0.007***
	Error	19	37493.96	1973.37		
				Within-Subject		
	IRA Task	1	6063.48	6063.48	6.88	**0.017***
	IRA Task * Order of IRA Experience	1	389.79	389.79	0.44	0.514
	IRA Task * Error Correction Group	1	82.73	82.73	0.09	0.763
	IRA Task * Order of IRA Experience * Error Correction Group	1	546.07	546.07	0.620	0.441
	Error (IRA Task)	19	16740.05	1401.75		
	Session	3	4205.24	1258.95	3.98	**0.012***
	Session * Order of Experience	3	3776.86	286.49	3.58	**0.019***
	Session * Error Correction Group	3	859.46	650.07	0.81	0.491
	Session * Order of Experience * Error Correction Group	3	1950.22	351.92	1.85	0.149
	Error (Session)	57	20059.66			
	IRA Task * Session	3	3028.35	1009.45	3.19	**0.030***
	IRA Task * Session * Order of Experience	3	4084.18	1361.39	4.32	**0.008***
	IRA Task * Session * Error Correction Group	3	207.94	69.31	0.220	0.882
	IRA Task * Session * Order of Experience * Error Correction Group	3	92.61	30.87	0.098	0.961
	Error (IRA Task * Session)	57	17985.11	315.53		
				Between-Subject		
Chain Length	Order of IRA Experience	1	1.94	1.94	0.42	0.525
	Error Correction Group	1	6.36	6.36	1.38	0.255
	Order of IRA Experience * Error Correction Group	1	20.61	20.61	4.46	**0.048***
	Error	19	87.74	4.62		
				Within-Subject		
	IRA Task	1	7.25	7.25	2.91	0.155
	IRA Task * Order of IRA Experience	1	15.67	15.67	4.74	**0.042***
	IRA Task * Error Correction Group	1	0.60	0.60	0.18	0.674
	IRA Task * Order of IRA Experience * Error Correction Group	1	2.30	2.30	0.69	0.415
	Error (IRA Task)	19	62.86	3.31		
	Session	3	12.29	4.10	4.51	**0.007***
	Session * Order of Experience	3	2.03	0.68	0.75	0.529
	Session * Error Correction Group	3	5.92	1.97	2.17	0.101
	Session * Order of Experience * Error Correction Group	3	2.85	0.95	1.04	0.380
	Error (Session)	57	51.82	0.91		

Table A11 Continued		df	SS	MS	F	P
	IRA Task * Session	3	7.59	2.53	2.58	0.063
	IRA Task * Session * Order of Experience	3	4.21	1.44	1.47	0.233
	IRA Task * Session * Error Correction Group	3	0.52	0.17	0.18	0.912
	IRA Task * Session * Order of Experience * Error Correction Group	3	0.51	0.98	0.17	0.914
	Error (IRA Task * Session)	57	55.96			
		Between-Subject				
PQ	Order of IRA Experience	1	0.66	0.66	0.55	0.468
	Error Correction Group	1	1.62	1.62	1.35	0.260
	Order of IRA Experience * Error Correction Group	1	6.40	6.40	5.32	**0.033***
	Error	19	22.84	1.10		
		Within-Subject				
	IRA Task	1	0.68	0.68	0.87	0.362
	IRA Task * Order of IRA Experience	1	3.14	3.14	4.01	0.060
	IRA Task * Error Correction Group	1	0.54	0.544	0.70	0.415
	IRA Task * Order of IRA Experience * Error Correction Group	1	0.17	0.17	0.21	0.650
	Error (IRA Task)	19	14.88	0.78		
	Session	3	3.86	1.29	6.85	**0.001***
	Session * Order of Experience	3	0.36	0.12	0.63	0.598
	Session * Error Correction Group	3	1.25	0.42	2.21	0.097
	Session * Order of Experience * Error Correction Group	3	0.65	0.22	1.15	0.337
	Error (Session)	57	10.71	0.19		
	IRA Task * Session	3	1.86	0.62	2.49	0.069
	IRA Task * Session * Order of Experience	3	1.29	0.43	1.73	0.171
	IRA Task * Session * Error Correction Group	3	0.27	0.09	0.37	0.779
	Interaction IRA Task * Session * Order of Experience * Error Correction Group	3	0.06	0.02	0.08	0.969
	Error (IRA Task * Session)	57	14.17	0.25		

Note: * Indicates $p < .05$ ** Indicates $p < .01$

Table A12
Examination of the Error Correction Procedure on the Individual IRA Tasks

		Forward Chaining IRA Task					*Backward Chaining IRA Task*				
IRA Variable	Source	Df	SS	MS	F	p	*df*	*SS*	*MS*	*F*	*p*
		Between-Subject					*Between-Subject*				
Response Total	Error Correction Group	1	946.91	946.91	0.297	0.592	*1*	*1143.96*	*1143.96*	*0.67*	*0.421*
	Order of IRA Experience	1	591.90	591.90	0.186	0.671	*1*	*222.08*	*222.08*	*0.13*	*0.721*
	Error Correction Group * Order of IRA Exp.	**1**	**13924.82**	**13924.82**	**4.3656**	**0.049**	*1*	*1098.66*	*1098.66*	*0.64*	*0.430*
	Error	21	66989.17	3189.96			*23*	*39212.15*	*1704.88*		
		Within-Subject					*Within-Subject*				
	Session	**3**	**6981.60**	**2327.20**	**3.95**	**0.012***	*3*	*790.60*	*263.53*	*1.42*	*0.245*
	Session * Error Correction Group	3	840.89	280.30	0.48	0.700	*3*	*688.72*	*229.57*	*1.23*	*0.304*
	Session * Order of IRA Experience	**3**	**7866.18**	**2622.06**	**4.45**	**0.007***	*3*	*116.10*	*38.70*	*0.21*	*0.891*
	Session * Error Corr. Group * Order of IRA	3	1315.50	438.50	0.74	0.530	*3*	*762.94*	*254.31*	*1.37*	*0.260*
	Error (Session)	63	37125.29	589.29			*69*	*12836.04*	*186.03*		
		Between-Subject					*Between-Subject*				
Chain Length	Error Correction Group	1	1.58	1.58	0.19	0.667	*1*	*0.02*	*0.02*	*0.00*	*0.949*
	Order of IRA Experience	1	7.56	7.56	0.92	0.350	*1*	*3.12*	*3.12*	*0.90*	*0.354*
	Error Correction Group * Order of IRA Exp.	1	11.20	11.20	1.36	0.257	*1*	*7.06*	*7.06*	*2.03*	*0.168*
	Error	21	173.36	8.26			*23*	*80.16*	*3.49*		
		Within-Subject					*Within-Subject*				
	Session	**3**	**19.83**	**6.61**	**4.23**	**0.009***	*3*	*2.319*	*0.77*	*1.42*	*0.246*
	Session * Error Correction Group	3	1.55	0.52	0.33	0.804	***3***	***5.76***	***1.92***	***3.51***	***0.020***
	Session * Order of IRA Experience	3	3.78	1.26	0.81	0.495	*3*	*2.091*	*0.70*	*1.28*	*0.289*
	Session * Error Corr. Group * Order of IRA	3	2.85	0.95	0.61	0.613	*3*	*0.44*	*0.15*	*0.27*	*0.847*
	Error (Session)	63	98.44	1.56			*69*	*37.69*	*0.55*		
		Between-Subject					*Between-Subject*				
PQ	Error Correction Group	1	0.79	0.79	0.37	0.549	*1*	*0.03*	*0.03*	*0.032*	*0.859*
	Order of IRA Experience	1	1.70	1.70	0.80	0.382	*1*	*0.58*	*0.59*	*0.69*	*0.414*
	Error Correction Group * Order of IRA Exp.	1	4.49	4.49	2.104	0.162	*1*	*1.54*	*1.54*	*1.85*	*0.187*
	Error	21	44.81	2.134			*23*	*19.18*	*0.83*		
		Within-Subject					*Within-Subject*				
	Session	**3**	**6.22**	**2.07**	**5.76**	**0.002***	***3***	***1.36***	***0.45***	***2.80***	***0.046****
	Session * Error Correction Group	3	0.37	0.12	0.34	0.794	*3*	*1.26*	*0.42*	*2.59*	*0.060*
	Session * Chain Type First	3	1.38	0.46	1.28	0.289	*3*	*0.64*	*0.21*	*1.31*	*0.278*
	Session * Error Corr. Group * Order of IRA	3	0.83	0.28	0.77	0.517	*3*	*0.15*	*0.05*	*0.31*	*0.820*
	Error (Session)	63	22.70	0.36			*69*	*11.16*	*0.16*		

Note: Forward and Backward Chaining IRA are not directly compared in this set of RMANOVAs. * Indicates $p < .05$ ** Indicates $p < .01$

Table A13

Description of the difference between Session 1 and 3, and between Session 5 and 7.

Order of IRA tasks	Forward Chaining IRA Task					Backward Chaining IRA Task				
	n	M	SD	Min	Max	n	M	SD	Min	Max
Forward Chaining IRA First										
Response Total	**17**	**18.02**	**37.99**	**-37.00**	**110.00**	16	-7.94	15.18	-24.00	34.00
Chain Length	**17**	**1.59**	**1.73**	**0.00**	**6.00**	16	0.19	1.17	-2.00	3.00
PQ	**17**	**0.78**	**0.99**	**-0.40**	**3.56**	16	0.13	0.62	-1.04	1.57
Backward Chaining IRA First										
Response Total	9	6.00	27.68	-51.00	39.00	**12**	**0.33**	**20.84**	**-39.00**	**33.00**
Chain Length	9	0.89	0.93	0.00	2.00	**12**	**0.33**	**1.30**	**-2.00**	**3.00**
PQ	9	0.61	0.71	-0.36	1.76	**12**	**0.36**	**0.69**	**-0.63**	**1.88**

Notes: Results from the first IRA task encountered are shown in bold, the Forward Chaining IRA task is on the left, and the Backward Chaining IRA task is on the right.

Table A14

RMANOVA of the difference between Session 1 and 3 when Error Correction Groups are Combined.

IRA Variable	Source	Df	SS	MS	F	p
		Between-Subject				
Response Total	Order of IRA Experience	1	4.30	4.30	0.005	0.946
	Error	23	20832.19	905.75		
		Within-Subject				
	IRA Task	1	2045.84	2045.84	6.96	0.015*
	IRA Task * Order of IRA Experience	1	545.58	545.58	1.86	0.186
	Error	23	6762.05	294.00		
		Between-Subject				
Chain Length	Order of IRA Experience	1	0.72	0.72	0.44	0.516
	Error	23	38.00	1.65		
		Within-Subject				
	IRA Task	1	10.43	10.43	8.67	0.007*
	IRA Task * Order of IRA Experience	1	0.35	0.35	0.29	0.596
	Error	23	27.65	1.20		
		Between-Subject				
PQ	Order of IRA Experience	1	0.04	0.04	0.093	0.763
	Error	23	10.25	0.45		
		Within-Subject				
	IRA Task	1	2.05	2.05	4.92	0.037*
	IRA Task * Order of IRA Experience	1	0.04	0.04	0.09	0.763
	Error	23	9.56	0.42		

Notes: “Difference” is a raw value from: “Value at the End (Session 3)” – “Value at the Beginning (Session 1)”
* Indicates $p < .05$ ** Indicates $p < .01$

Table A15
Correlations between Forward and Backward Chaining IRA Measures

	N	r	P
Session 1 & 5 Combined			
Response Total	51	0.264	0.061
Error Total	51	0.253	0.073
Chain Length	51	0.055	0.703
Accuracy	51	0.031	0.829
PQ	51	0.052	0.719
Difference Value Session 1- 3 (or 5-7)			
Response Total	25	0.539**	0.005
Chain Length	25	0.161	0.442
PQ	25	0.035	0.867

Notes:
* Indicates $p < .05$
** Indicates $p < .01$

Table A16
Description of WPPSI-III scores and

	N	% of Sample	Mean	SD	P
Complete WPPSI-III Sample					
Composite Score Verbal	46	83%	103.26	13.29	
Composite Score Performance	46	83%	100.43	13.56	
Composite Score Full Scale	46	83%	103.22	12.43	
Gender and Verbal IQ					0.030
Male Verbal IQ	25	45%	99.4	11.52	
Female Verbal IQ	21	38%	107.85	14.04	
Research Site and Verbal IQ					0.027
Location A	6	10%	101.5	8.98	
Location B	13	23%	113.15	13.53	
Location C	5	9%	101.8	2.86	
Location D	10	18%	95.7	10.25	
Location E	8	14%	102.5	12.65	
Location F	4	7%	96	19.61	
Error Correction Group and Verbal IQ					0.022
Group 1 – Automatic Error Correction	15	27%	107.27	14.27	
Group 2 – Contingent Error Correction	8	14%	110.625	9.91	
Group 3 – 2 Day, No Error Correction	23	41%	98.09	12.06	

Notes:

Table A17
Description of DCCS measures and correlations with IQ.

	n	Mean	SD	Min.	Max.	Full Scale IQ	Verbal IQ	Performance IQ
						n = 46	n = 46	n = 46
Phase	53	2	0.97	0	3	-0.031	-0.006	0.053
Correct- Phase-Criteria	53	12.28	6.67	0	23	-0.113	0.053	-0.019

Notes:
* Indicates *p* < .05
** Indicates *p* < .01

Table A18
Description of Session 1 (combined with session 5) IRA Task Performance

Measure	Mean	SD	N	Minimum	Maximum	P
Forward Chaining IRA Task						
Response Total	70.1	21.19	52	31	124	
Error total	41.40	11.49	52	16	69	
Chain Length	2.4	1.22	52	1	6	
Accuracy	0.38	0.15	52	0.21	0.80	
PQ	1.46	0.60	52	1.0	3.24	
Backward Chaining IRA Task						
Response Total	64.92	18.72	53	30	124	
Error total	42.08	13.17	53	20	78	
Chain Length	2.30	0.99	53	1	6	
Accuracy	0.34	0.14	53	0.06	0.79	
PQ	1.39	0.51	53	1.00	3.57	
Backward Chaining IRA Accuracy and Language Exposure						0.027
English Only	0.32	0.13	43	0.06	0.79	
English and Spanish	0.46	0.13	8	0.32	0.62	
English and Other	0.28	0.01	2	0.28	0.29	
Backward Chaining IRA PQ and Language Exposure						0.017
English Only	1.30	0.48	43	1.00	3.57	
English and Spanish	1.85	0.51	8	1.38	2.62	
English and Other	1.38	0.18	2	1.25	1.50	

Notes: PQ Max Possible for this Task = 5.0

Table A19
Correlations between Foreword Chaining IRA Task Session 3 (combined with session 7) and IRA Difference Values with age, IQ, and DCCS.

	Age in Months	Full Scale IQ Composite Score	Verbal IQ Composite Score	Performance IQ Composite Score	DCCS Phase	DCCS Correct
Session 3						
Response Total	0.645**	-0.019	-0.544**	0.394	0.400*	0.394*
Error Total	0.166	0.066	-0.025	0.182	0.064	0.061
Chain Length	0.558**	0.003	-0.447*	0.336	0.329	0.357
Accuracy	0.378*	0.032	-0.362	0.255	0.263	0.329
PQ	0.566**	0.007	-0.436*	0.274	0.393*	0.422*
Difference Value						
Response Total	0.716**	-0.091	-0.537*	0.318	0.317	0.306
Chain Length	0.601**	0.084	-0.289	0.288	0.262	0.290
PQ	0.540**	-0.014	-0.322	0.072	0.290	0.322

Notes: Age n = 26, IQ n = 21, DCCS n = 26
* Indicates $p < .05$
** Indicates $p < .01$

Table A20
Correlations between Backward Chaining IRA Task Session 3 (combined with session 7) and IRA Difference Value with age, IQ, and DCCS.

	Age in Months	Full Scale IQ Composite Score	Verbal IQ Composite Score	Performance IQ Composite Score	DCCS Phase	DCCS Correct
Session 3						
Response Total	0.643**	-0.123	-0.460*	0.182	0.438*	0.379*
Error Total	0.303	-0.266	-0.195	-0.323	0.361	0.368*
Chain Length	0.227	0.017	-0.285	0.330	-0.030	-0.125
Accuracy	0.328	0.215	-0.178	0.529**	0.129	0.080
PQ	0.330	0.024	-0.337	0.399	0.099	-0.018
Difference Value						
Response Total	0.490**	-0.414*	-0.636**	-0.076	0.409*	0.378*
Chain Length	0.242	0.009	-0.250	0.373	0.089	-0.003
PQ	0.278	-0.009	-0.312	0.348	0.174	0.075

Notes: Age n = 29 , IQ n = 24, DCCS n = 29
* Indicates $p < .05$
** Indicates $p < .01$

Table A21
Standard Regression Analysis

	B	Standard Error	β	R^2
	Log10 Transformation Forward Chaining IRA Error Total			
	F(2, 41) = 1.120 p = 0.336			
Full Scale IQ Composite Score	0.001	0.002	0.076	0.052
DCCS Level	0.031	0.023	0.206	
	Log10 Transformation Backward Chaining IRA Error Total			
	F (2, 43) = .548 p = .582			
Full Scale IQ Composite Score	0.000	0.001	-0.044	0.025
DCCS Level	-0.019	0.020	-0.148	

Note:

Appendix B

Instructions from Paule 1988b

"You will notice that there are four levers at the bottom of the panel. Above the levers are six colored lights and above these are a correct answer light with a happy face and an incorrect answer light. When the game begins the levers will come out of the machine and the first colored light will light. The light tells you that you only have to press one lever to receive a nickel. You may push any lever that you want. If it is the correct lever the happy face light will light and you will receive a nickel. If it is the wrong lever the incorrect light will light and you can try again. Once you learn which lever is correct you should remember it because it will always be the correct lever. You can continue pressing the lever and receiving nickels until there is a waiting period and the second colored light will then light.

You must then try and find which lever matches with the second colored light. It can be any of the four levers. When you find the correct lever the happy face light will light and the second colored light will go out. The first colored light will again light. You can now press the lever that goes with the first light and receive a nickel. When you find out which lever goes to the second colored light, remember it because it will always go to that light. Although you should try to do this quickly in order to get more nickels you should not go faster than the lights. After you push a lever wait until the happy face light or incorrect answer light goes out and a colored light comes on before you push another lever. After you play the game with two levers there will be a waiting period and the third colored light will light. Now you must press three levers to get a nickel. Remember the levers you learn will always be the same for today's game. The game will build up to four, five and six levers needed to receive a nickel".

Instructions from Baldwin et al. 2012

"Notice on the machine in front of you that there are four levers, here," [narrator points to the four levers] "at the bottom of the panel. Above these levers are six colored lights, here" [points to the six lights]. "Above these are a correct-answer light, here,"[points to the answer light on the left] "with the happy face and an incorrect-answer light, here" [points to the answer light on the right]. "When this game begins, all four levers, here," [points to the four levers along the bottom] "will come out of the machine. You will also notice this light, here," [points to colored light that is farthest to the right] "will light" [the four levers come out of the machine and the colored light on the far right lights up]. "This light" [points to the illuminated (red) light on the far right that is lit] "tells you that you only have to press one lever to receive a nickel. You may push any lever that you want. If it is the correct lever, the happy face light, here," [points to the answer light on the left] "will light and you will receive a nickel. If it is the wrong lever, the incorrect answer light, here," [points to the answer light on the right] "will light and you can try again. Let's try and see if we can find out which lever it is." [The narrator presses the lever on the far right and points to the incorrect-answer light that is now lit]. "No, that's incorrect." [The narrator presses the second lever from the right and points to the incorrect-answer light that is now lit]. "That one is incorrect." [The narrator presses the third lever from the right and points to the incorrect-answer light that is now lit]. "That one's incorrect. We now know that it must be this lever here," [points to the lever on the far left and presses it; the correct-answer light is illuminated and the narrator receives a nickel]. "Yes" [points to the correct-answer light]. "The happy face light lit and we received a nickel. Once you learn which lever is correct, you should remember it because it will always be the correct lever to play on this game on this day. You can continue pressing the lever and receiving nickels until there's a waiting period" [the narrator presses far left lever again and receives a nickel]. "See, we continue to get nickels" [narrator presses the lever again and receives another nickel, after which the colored light on the far right goes out]. “Now there’s a waiting period.” [The second colored (green) light from the right lights up]. “We see that the second light, here,” [points to the green light] “has lit. You must now find which lever matches the second colored light” [points to the light again]. “When you find which one matches, you can press that lever and the second colored light will go out. The first light” [points to the red light farthest to the right] “will light and you can go back and press the lever that goes with the first light to receive a nickel. Let’s try to find out which lever it is.” [The narrator presses the lever on the far right and points to the incorrect-answer light that is now lit]. “No, that’s incorrect.” [The narrator presses the second lever from the right and points to the incorrect-answer light that is now lit]. “No, that’s incorrect.” [The narrator presses the third lever from the right and points to the correct-answer light that is now lit]. “Yes, that was the correct one. We saw the happy face light light and then we went back to the first light” [points to the red light on the far right, which is now the only one lit]. “You now go back to the first lever and you’ll receive a nickel” [narrator presses the lever on the far left and receives a nickel and the green light is re-illuminated]. “See, now all you have to do is repeat the process” [points to the second (green) colored light]. “Here,” [presses the third lever from the right] “back to the first light” [points to the red light] “and here” [presses the lever on the far left and receives a nickel]. “After you push a lever, wait until the happy face light” [points to the answer light on the left] “or the incorrect-answer light” [points to the answer light on the right] “goes out and the lever light lights again first before you start pressing another lever, like this” [the narrator presses the third lever from the right and points to the correct-answer light]. “Wait until this light goes out” [the correct-answer light goes out and the narrator points to the red colored light on the far right] “and this one lights back before you press the next lever” [presses the lever on the far left and receives a nickel] “to receive your nickel.” [The red colored light goes dark]. “See, now there’s another waiting period and the third light lights up” [points to the third (orange) light from the right]. “The game will continue building to four” [points to the fourth (blue) light from the right], “five” [points to the fifth (yellow) light from the right], “and six levers” [points to the light (pink) on the far left]. “Remember the sequence and you’ll continue getting nickels.”

Instructions without description of Narrator.

“Notice on the machine in front of you that there are four levers, here, at the bottom of the panel. Above these levers are six colored lights, here. Above these are a correct-answer light, here, with the happy face and an incorrect-answer light, here. When this game begins, all four levers, here, will come out of the machine. You will also notice this light, here, will light. This light tells you that you only have to press one lever to receive a nickel. You may push any lever that you want. If it is the correct lever, the happy face light, here, will light and you will receive a nickel. If it is the wrong lever, the incorrect answer light, here, will light and you can try again. Let’s try and see if we can find out which lever it is. No, that’s incorrect. That one is incorrect. That one’s incorrect. We now know that it must be this lever here, Yes. The happy face light lit and we received a nickel. Once you learn which lever is correct, you should remember it because it will always be the correct lever to play on this game on this day. You can continue pressing the lever and receiving nickels until there’s a waiting period. See, we continue to get nickels. Now there’s a waiting period. We see that the second light, here, has lit. You must now find which lever matches the second colored light. When you find which one matches, you can press that lever and the second colored light will go out. The first light will light and you can go back and press the lever that goes with the first light to receive a nickel. Let’s try to find out which lever it is. No, that’s incorrect. No, that’s incorrect. Yes, that was the correct one. We saw the happy face light light and then we went back to the first light. You now go back to the first lever and you’ll receive a nickel. See, now all you have to do is repeat the process. Here, back to the first light and here. After you push a lever, wait until the happy face light or the incorrect-answer light goes out and the lever light lights again first before you start pressing another lever, like this. Wait until this light goes out and this one lights back before you press the next lever to receive your nickel. See, now there's another waiting period and the third light lights up. The game will continue building to four, five, and six levers. Remember the sequence and you'll continue getting nickels."

440 words – Reading Level – Grade Level 4.2

Appendix C

Supplemental Literature Review

Appendix C

Part 1: Early Development in the Animal Lab

Repeated acquisition was first described by Boren (1963) at an APA convention and first published in Boren and Devine (1968). The original procedure used a complicated shaping program to generate performance of behavioral chains. Animals (rhesus monkeys) received a food pellet for a response on any one of 12 levers. These 12 levers were organized in a single line with four groups of three levers, investigators numbered these levers 1-12 from left to right for convenience. When training was completed, all behavioral chains involved a response on one of the levers in each of the four groups and the final behavioral chain (four responses) always moved from left to right across the groups of levers. A light above each group of levers lit up above when the chain incremented and a response on that group of levers was required. The second step in the shaping procedure was to reinforce a response on any of the far right three levers (#10,11 or 12). Next a two-response chain was required, and any of the third group of levers (#7,8 or 9) could be pressed followed by any of the far right, fourth group of three levers (#10,11,12). This procedure extended until the animals were completing a four response chain across all four groups of levers. No specific lever, or sequence of levers, was considered correct. Following this, the animals were placed on an FR2 schedule and then an FR5 schedule. At this point a 15 second time-out was added for lever presses that were incorrect (pressing a lever not in the correct group). Finally responses on only specified single levers from each set of levers were reinforced. A last step was to vary a single element of the learned behavioral chain. Finally the investigators could present the animals with a new four link sequence during each session and measure the acquisition across conditions. During these sessions a four response behavioral chain was always required to obtain reinforcement. The measurements used included the total number of errors, and *performance errors* defined as errors made after the sequence had been established during a session (last 65 reinforcers of a session), while *learning errors* were those that occurred before this mark. Average errors were used to compare training manipulations across single animals.

The first important modification of Boren and Devine (1968) was by Thompson (1970) who simplified the shaping procedure in order to establish the behavior chain. Working with pigeons Thompson used only 4 response keys and dropped the left to right sequence requirements. During each session a four-link chain was required to earn reinforcement (following magazine training and the training of a two-link chain), after each response the key lights changed such that the animals experienced yellow key lights, following a correct response green key lights, following a correct response red key lights, following a correct response white key lights, following a correct response five second access to mixed grain as a reinforcer. When stable responding was established Thompson conducted what he termed “Tandem” probes. In these probes the coloring of the key light was removed and between each correct response within a sequence the key lights, now white, dimmed for a brief period of time.

Pieper (1976) describes a second important modification of Boren and Devine’s 1968 task. In this modification, again with nonhuman primates, six horizontally mounted response levers were used. Located above each lever was a circular disk light with the ability to vary illumination level. These light cues were used as an initial signal of which response was required, the light above the required lever lit up at full brightness. Then throughout the session the light was faded out, growing dimmer after each correct response. The technology allowed for 8 levels of illumination between full illumination and fully dark. The final result was a tandem schedule. Pieper also introduced the critical modification that differentiates Repeated Acquisition and Incremental Repeated Acquisition: the development of the behavioral chain across the session. Animals began with a one-link chain, when the animal could execute this sequence without the cue light the chain incremented to a two response chain that used the cue lights. This modification enhanced the behavioral control exerted by the stimuli within the chain schedule; by establishing responding in what could be described as an attempt toward *errorless learning* and fading these stimuli out until the animal was responding on a chain schedule. The maximum chain length was a six response chain. Daily sessions lasted for 50 minutes or until the animal had successfully completed a six lever sequence. Pieper started animals in a learning phase where each day a new sequence was learned. The primary dependent measure utilized by Pieper was the maximum sequence length completed during the 50 minute session. Unfortunately detailed results were not reported. Pieper did not describe the incrementing method in detail so it is unclear if forward or backward chaining methods were used.

Procedural Issues

Advancement Criteria. Another procedural development that has seen little discussion is the refinement of criteria for advancing through a chain. The criteria set by the earliest IRA studies were different from more current studies. Pieper (1976) incremented the sequence based on successive correct responses within an eight level fading procedure that was sensitive to errors, and thus required an inconsistent number of correct responses for each individual animal. In contrast Weinberg and Killam (1978) set consistent and strenuous within-session incrementing criteria, 21 successive errorless chains. Later work by Paule and McMillian (1984) with rats set equally stringent within-session advancement criteria, 40 correct responses, although consecutive errorless sequences were not required. This can be contrasted with more recent IRA work with rats and mice (Bailey et al. 2010, Spencer & Newland, 2010a; Johnson et al, 2010) which has used lower consecutive errorless sequence requirements that vary at each different link of the chain, for instance ten consecutive correct responses on the first link, five on the second and third link and so on. An important difference between early and later work that might interact with this procedural condition is the treatment of errors; in later work errors reset the chain and this may moderate the need for a stringent within session advancement criteria.

Use of both performance and learning components Cont. Detailed information on the impact of a decision to present various components in a multiple schedule or during separate sessions are not available. At least one study (Bailey et al. 2010) suggests that separate sessions might contribute to animals attaining longer than previously published chains. The use of a performance and learning component is one procedural element of the IRA task that has seen a subtle shift in implementation. Later studies (Bailey et al. 2010; Spencer & Newland, 2010a, Johnson, Bailey, Johnson & Newland, 2010) use a performance phase as an initial training and baseline phase, this phase always comes first in these studies. Various criteria are used to indicate stable performance at which time a second baseline learning condition is introduced. Finally drug challenges are introduced and the conditions and sequences are arranged based on pharmacological considerations.

The distinction is particularly useful during drug challenges allowing the selective effects of a substance to be determined. These components can be combined into multiple schedules, or used during different sessions according to the needs of the investigator. In some cases the learning component and the performance sequence are differentiated from each other via specific stimuli; red or green backgrounds (Thompson and Moerschbaucher, 1979) or the presence or absence of olfactory stimuli (unpublished work; presented as Spencer & Newland, 2010a).

Stimuli Arrangements to Enhance Responding. These are not the only examples of procedures which used various stimuli as prompts to more specifically guide behavior (Pieper, 1976). The arrangement of these prompting stimuli may affect the schedule such that it is no longer a chain schedule (ie all stimuli are faded out so that what remains is a tandem schedule; Pieper, 1976). Even when the schedule remains a chain schedule the specificity of a chain schedule which prompts the correct response, even when this element is faded out, must be viewed as extremely different from a chain schedule which only cues another response.

Treatment of Errors. Another alternative in the treatment of errors is presented in a study by Cohn, Cox and Cory-Slechta (1993). In this study an error increased the ratio of reinforcement schedule from a continuous reinforcement schedule to an FR2, increasing the cost of an error. This manipulation has only been used once in a repeated acquisition procedure but appears to increase the difficulty for the animals, because of this manipulation the authors reported using an additional criteria to make experimental decisions.

Chain Structure. The structure of the behavioral chain in relation to the response keys has frequently been cited as an important variable but less frequently been manipulated as a variable of interest. Early work by (Boren and Devine, 1968) included a consistent criterion for chain selection that required adjacent responses on each successive group of levers. Other work explicitly avoided adjacent responses on successive links of the chain (Thompson, 1970; Pieper, 1976). In all studies the various chains were carefully selected to be equivalent and typically avoided the successive repetition of a response.

Thompson (1973) used the following criteria for chain selection, which has been adopted by others (Zayac & Johnston, 2008).

> "First a correct color position in one session was not repeated in the following session. Second, simple orders, such as the left key in each group of colors, were avoided; adjacent positions in each sequence, were always different, although each position occurred at least once. Third, within a set of six sequences, each key position appeared equally often (twice) in each color" (p. 507-508).

Work by Lattal and Crawford-Goodbey (1985) also detected improved performance of homogenous chains, chains in which the behavioral components were similar, compared to heterogeneous chains

Appendix C

Part 2: Additional information from studies conducted by the NCTR with humans.

The first use of the OTB with humans by Paule and colleagues (1988b) sampled a group of twenty children between the ages of 3 and 11 years. The group purposefully included children with 'learning disabilities' and ADD in an effort ensure they sampled a wide range of participant abilities and to explore if the OTB component tasks could discriminate clinical differences. In all OTB validation studies the same panel originally designed for Rhesus monkeys was used. This was a large operant instrument panel with three 'press plates' (flat response instruments), four retractable levers and various lights and speakers. The panel was installed in a research room at a local hospital and was modified to dispense nickels to children instead of the food pellets used in the animal lab. The IRA component lasted for fifteen minutes, used backward chaining methods to build the behavioral chain, and allowed participants to attain a maximum six link chain. Each time a participant met criterion, a 30 second blackout preceded the lengthening of the chain. The four retractable levers were used, and were configured in a single row. The criterion for incrementing between chain links was three correct responses and a participant could earn a maximum of 18 reinforcers. The description of the task is not totally clear but it appears that participants were stopped when they reached 18 reinforcers, regardless of whether they still had time left or how far they had gotten through the task.

A book chapter (Paule & Cranmer, 1990) detailed additional information from this original group of participants. Learning curves (cumulative records) were created for each of the four 6-year-old participants. Only one of these participants was able to complete the entire six lever sequence in the 15 minutes provided at the mastery level, a typically-developing child included as a member of the control group. The chapter describes this participant as completing the six-lever level. Alternatively, another participant is described as not demonstrating mastery at the six-lever level, while another did not demonstrate mastery at the five lever level. Finally the fourth participant did not make it through the two lever level. Based on this chapter it is clear that some participants completed the OTB component tasks more than once, but no description is given of the procedures for doing this. For instance the time between testing (described as a few months) is not detailed, nor is it clear if all participants completed the test a second time. The general conclusion based on these results was that ADD and LD children do not respond the same way as the children selected in the control group on the IRA task. Interestingly, the investigators describe their procedure as all nonverbal, with the exception of the instructions. The instructions were designed as "partially verbal in nature" but investigators note that performance differences might be traced to problems with understanding the instructions.

A second study by Paule, Forrester, Maher, Cranmer & Allen published in 1990(b) compared OTB performance in children with monkeys at the NCTR lab. Seventy one children between 4 and 8 years of age completed the IRA task with mean accuracy by age ranging from 33%-76%. The NCTR investigators noted the steady improvement across age and speculated that the improvement would plateau at about 7 or 8 years. Interestingly, response rate on the IRA task was the only OTB variable for which significant differences between monkey and child performance could be detected, with children responding, regardless of age, around 0.5 responses per second and monkeys responding well over 1 response per second. Since the same device was used the investigators indicated this difference probably reflected the high level of training achieved with the monkeys and the effectiveness of the reinforcement schedule used with these animals.

Appendix C

Part 3: Direct comparisons of Forward and Backward chaining.

The studies included in this analysis were first gleaned from other work included in this review that discussed forward and backward chaining (Bailey et al. 2010, Johnson et al. 2010, Batra & Batra 2006) as well as from the reference section of an ABA text book (Martin & Pear, 2007) and finally from a search of PsychInfo using the key terms "Forward chaining AND Backward chaining". From this list of sources books and studies published only in dissertation abstracts were omitted, this yielded a total of 13 direct comparisons of forward and backward chaining. This analysis is not exhaustive, other studies on the topic may exist. It is also important to note that many of these studies also include a comparison to whole task or total task chaining. Information regarding these comparisons is not included in the analysis; however, it is notable that there was not a consensus on the superiority or inferiority of whole task chaining methods either. Studies which only compared forward chaining to whole task chaining or only compared backward chaining to whole task chaining were also omitted. These 13 studies were analyzed based on outcome measures and participant populations

When participant was considered these 13 studies were broken down into three categories; special needs populations, typically functioning adults (college students or military personnel), and a single animal study. Five studies focused on special needs populations; and five of the outcomes indicated no difference between forward and backward chaining, while a single outcome indicated backward chaining was superior to forward chaining. Six studies focused on college students or military personnel; three outcomes suggested there was no difference, five outcomes suggested forward chaining was superior. The single animal study indicated backward chaining was superior. This analysis yielded the most clear cut results in terms of practical advice for future investigators; for typically functioning participants (such as college students or military personnel) it is probably preferable to use forward chaining. In contrast when working with special needs populations there isn't (based on these 13 studies) strong evidence to indicate the use of either chaining method over the other, and so the determination must be based on something else. Slocum and Tiger (2011) explicitly attempted to use short forward and backward chaining procedure to predict how children with DD would perform on a longer chain task. In this study investigators did not find that children consistently performed better under one chaining type; children who performed better on the short forward chaining task did not also perform better on the long forward chaining task.

The examination of response chain type across these studies was more difficult given the variety of tasks. The largest category was applied tasks; these seven studies included response chains ranging from assembly of a bicycle brake to playing a musical instrument (a keyboard) and military procedural tasks (missile preparation). This was the only category where backward chaining ever appeared to be superior to forward chaining, and then only on two occasions. For six different task outcomes with an applied task there was no difference between chaining methods, and once forward chaining was superior. The remaining studies were broken down into two other categories; four studies used an arbitrary response pattern (for instance arbitrary response buttons, or motor movements) and two studies used language or math based tasks. The studies using an arbitrary response indicated no difference for two outcomes and forward chaining for two outcomes. The studies using language or math tasks indicated two outcomes indicating no difference and two outcomes indicating forward chaining. Once again empirical evidence does not support a strong preference for forward chaining or backward chaining. Text books on application suggest the decision between forward and backward chaining include a consideration of reinforcement available during training (Martin & Pear 2007). This consideration is particularly relevant to the applied task category where we can distinguish between *natural reinforcers* and reinforcers arranged solely for the purpose of training. In this case behavior taught under forward chaining is described as particularly sensitive to the reinforcer selected, as the natural reinforcer may not occur at all until the entire chain has been mastered. Placement and exposure to reinforcers during chaining procedures is a critical theoretical issue explored more below. It is important to remember the definition of a response chain as a temporally extended unit of behavior (Catania, 2007) would not distinguish between applied tasks with a natural reinforcer and seemingly arbitrary laboratory tasks (pressing buttons on a computer screen) where reinforcers are apparently equivalent in terms of relation to the environment. A more critical distinction is probably apparent in the transition from a procedure where chaining is used strictly for training to an IRA task. In this case chaining is no longer used solely to establish a response chain; instead chaining becomes the driving means of analyzing behavior repeatedly and it is the acquisition of the behavior, rather than the task itself that is meaningful.