Spring 2019

Self-Control Abilities in Squirrel Monkeys (Saimiri sciureus)

Renee Russell
Bucknell University, rcr024@bucknell.edu

Recommended Citation

Russell, Renee, "Self-Control Abilities in Squirrel Monkeys (Saimiri sciureus)" (2019). *Master's Theses*. 215.
https://digitalcommons.bucknell.edu/masters_theses/215

I, Renee Russell, do grant permission for my thesis to be copied.

SELF-CONTROL ABILITIES IN SQUIRREL MONKEYS (*SAIMIRI SCIUREUS*)

by

Renee C. Russell

A Thesis

Presented to the Faculty of

Bucknell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science in Animal Behavior

Approved: Peter Judge
Adviser

Peter Judge
Program Director

4/25/19
Date

Acknowledgements

There are many people to thank for helping me complete this journey. First and foremost, to Dr. Peter Judge, my thesis adviser, thank you for your never ending patience and guidance as I undertook this research and made a questionable choice about going to Costa Rica. Your expertise and continued willingness to provide feedback has been invaluable. I cannot thank you enough for all of the opportunities and memories over the past two years.

To my committee members, Dr. Reggie Gazes and Dr. Elizabeth Capaldi, thank you for sharing your time and your insight throughout this process – and an extra thank you to Reggie for your advice and infectious enthusiasm.

To our fabulous animal caretakers, especially Gretchen and Amber, thank you for not only providing the absolute best care for my research subjects but also for me and my sanity.

To Julie Berhane, I do not know what I would have done without you by my side throughout this entire process. Through every conference, every deadline, every study "party", and every (rare) occasion we got to be actual people, I was glad to have you with me. You are my academic other half…long live Renulie Berhussell.

To Casey Early, who may be the only other person crazy enough to think squirrel monkeys are great research subjects, thank you for convincing Peter to let you attempt token exchange training, you started it all. To Missy Painter, thank you for showing me the ropes and being my office buddy. To Sara Faiad, thank you for always letting me

vent, for always justifying my purchases, and for being one of my closest friends…UW is lucky to have you.

To all of my academic mentors and peers, past and present, I am forever grateful for the time and energy you invested into my education and my being. I am a better researcher – and person – because of you. To Dr. Shedd, thank you for inspiring my wonder and appreciation of all animals, your dedication to them and your students truly sets you apart. To Currier, you have an incredible knack for knowing exactly what I need, when I need it, whether it be an academic sounding board, a yogic perspective, or most importantly, a true friend. To Kati Biggs (bae), a proper thank you would be as long as this entire thesis, but you will have to settle for two sentences (including this one). Thank you for remembering more details about my life than I do, I would truly be lost without your friendship. To Averie Morgan, thank you for getting me through concussion land, and one of the hardest years of my life. To both of you and MacKenzi Brown, thanks for always being down to meet me in the Dell. To Melanie Freeman and Sara Woodward, thank you for continuing to be two of my loudest cheerleaders.

To my yoga friends, thank you for being the warmest support system I could ever want or need and for constant hugs and reminders that I am never alone on this wild journey. And finally, to my parents, thank you for your unwavering support, love, and sacrifices (most recently, braving the Pennsylvanian “Spring” to watch my thesis defense) throughout the past 25 years, I literally would not be here without you.

Table of Contents

Behavioral Inhibition Tasks (A-not-B, Cylinder)

Food Exchange Task

Accumulation Task

List of Tables

List of Figures

Abstract

Self-control has been studied extensively in both humans and nonhumans in relation to planning, goal-oriented behavior, and overall higher cognitive function. These investigations have resulted in a vast literature pool afflicted by differing definitions, procedural inconsistencies, and numerous paradigms that were thought to measure self-control. I utilized a within subject design to address the question of what these existing tasks were actually studying and if they produced similar results. The present study tested squirrel monkeys on four tasks. Two of which were widely accepted self-control paradigms (food exchange and accumulation) that involved refraining from choosing a low-value reward in favor of a high-value reward. Importantly, these two rewards differed qualitatively (choosing a more preferred item) or quantitatively (choosing a larger amount of the same item) depending on the task/phase. The other two tasks were highly contested self-control paradigms (cylinder and A-not-B) that involved the inhibition of a prepotent motor response. All squirrel monkeys tested were capable of displaying self-control on all tasks, to some extent. However, the results indicate that not all self-control paradigms measure the same aspect of self-control, as individual performances varied across the tasks. Task type and previous experience with other self-control tasks did not predict the variability in performance, but the qualitative or quantitative nature of the rewards did. Ultimately, this suggests that researchers should not compare the results from one task to that of another, especially if the nature of the reward differs. Finally, the present studies added to the existing literature pool of self-control studies, specifically addressing the paucity of data for New World monkey species in a way that allowed for the results to be directly compared to previous studies. Overall, the squirrel monkeys performed quite well on all tasks, indicating that they were promising subjects for future studies of self-control.

Introduction

Animals, including humans, may benefit from restricting their behavior in certain social situations or from making decisions that will either increase their yield or prevent misfortune in the future. Self-control, and its supposed opposite, impulsivity, have been studied extensively in both humans and nonhumans in relation to planning and goal-oriented behavior, substance abuse, and overall higher cognitive functions (Beran, 2018). Self-control has been investigated since the early 20th century, creating a vast literature pool afflicted by differing definitions and procedural inconsistencies (Eisenreich & Hayden, 2018). Historically, the behavioral definition of self-control was the act of choosing a higher valued, but more delayed reinforcer over a lower valued, less delayed reinforcer (Rachlin & Green, 1972). More recently, self-control has been viewed as an ability or skill that can be learned. Some researchers even suggest that it is a capacity that can be depleted with overuse or increased through training (e.g., Baumeister, Vohs, & Tice, 2007), though this hypothesis has not been supported in nonhuman primate literature (De Petrillo et al., 2015; Parrish, Emerson, Rossettie, & Beran, 2016; Parrish et al., 2018). Decades of self-control studies have utilized multitudes of paradigms, each striving to uncover which species demonstrate this ability and to what extent. The research has led to an active debate about what these paradigms are actually testing and how their results should be interpreted (for an overview, see: Beran, 2015; Beran, 2018; Eisenreich & Hayden, 2018; Flessert & Beran, 2018). However, there is a lack of clear, experimental evidence to corroborate the arguments that comprise the debate.

Delay of gratification is a form of self-control that is typically broken down into two components (Beran, 2018) and is considered to be a prerequisite for complex goal-directed action (Mischel, 1974). The first component, delay choice, is the initial decision to forgo an immediately available option in favor of a more valuable, but delayed option. The second component used to be known as "bridge the delay interval," but has since come to be known as delay maintenance as it refers to the continued decision to "maintain" that initial choice when continually faced with the immediate option (Beran, 2002). A popular example is one of going on a diet. If you decide one morning that you are going to start eating healthier, you are demonstrating delay choice. However, if later that night you are confronted with a delectable sweet treat, you could decide to stay true to your diet, thus exercising delay maintenance, or you could indulge yourself (and your impulsivity) and defect in regards to your choice to go on a diet. As demonstrated in this example, there is quite a difference between the two components, yet both are necessary to delay gratification long enough to reap the benefits of your self-control (increased health).

The majority of self-control tests fall into one of three categories: delay choice, delay maintenance, and hybrid delay. In delay-choice tasks, participants are able to choose between two reward options, typically thought of as a "small reward sooner" or a "large reward later," but rewards could differ qualitatively as well as quantitatively. Once the participant has made their choice, they cannot defect to the other option. For example, once a subject picks the more preferred but delayed option, they cannot change their mind and instead choose the less preferred, but immediate option. In this way, participants can choose to delay their reward in order to procure a better one, but do not have to maintain their choice through the delay period. In delay-maintenance tasks, participants must maintain their self-control throughout a delay period if they are to receive the better reward. In other words, they are able to defect and choose the immediate reward at any point during the delay period, effectively terminating the trial. Hybrid tasks typically incorporate a delay-maintenance task into an existing delay-choice paradigm, so that the participant must first chose the more delayed option and then refrain from defecting to the small, immediate reward throughout the course of the trial. Controversy surrounds a possible forth category of self-control tests, behavioral inhibition tasks, which typically involve a transparent barrier that the participant must maneuver around in order to obtain a reward. Some researchers (MacLean et al., 2014) have used these tasks as a measure of self-control in experiments aimed at elucidating the evolutionary origins of self-control. Other researchers (namely Beran, 2015) have openly challenged behavioral inhibition tasks, asserting that they are not accurate measures of self-control, and at best, offer a measure of motor-inhibition, which is in some ways related, but not synonymous to, self-control. Below, I review some of these paradigms and what they might reveal about self-control abilities in nonhumans.

Exchange tasks

Exchange tasks, at the foundational level, involve giving the subject an item of lower value (or an item associated with a lower value) that they can then exchange (either immediately or after a time delay) with an experimenter for a higher valued reward. The item can be a small, inedible object as in token exchange tasks or a food reward as in delayed exchange tasks, which I will refer to as food exchange tasks for clarity.

Token exchange tasks are sometimes referred to as bartering paradigms and are considered to be a delay-choice paradigm because once the token is chosen, they have forfeited the lower valued food. Token exchange has been used in a number of experiments to investigate a variety of topics (see Hackenberg, 2009, for a review), however, Beran and Evans (2012) and Judge and Essler (2013) simultaneously, but independently integrated the use of tokens into similar tests of self-control. Beran and Evans (2012) ultimately gave three language-trained chimpanzees (*Pan troglodytes*) a choice between an immediately available medium-value food and a lexigram token that they could exchange for its associated high-value food. Two of the three participants chose the high-value token significantly over chance, both when they could immediately exchange the token and when they had to wait two to three minutes to exchange.

Judge and Essler (2013) conducted a similar experiment with brown capuchin monkeys (*Sapajus apella*) using small pieces of hardware as tokens. After a participant had demonstrated their knowledge of the token associations (by choosing to exchange all of the high-value tokens available before the low-value tokens), they were given a series of “bartering up” test sessions. First, participants were given the low-value token to exchange with an experimenter. After a successful exchange, the experimenter provided them with a choice between the immediately consumable, low-value reward that was associated with that token or a token that was associated with a high-value reward that they could then exchange for the more preferred reward. Two of the four participants selected the high-value token over the low-value food significantly more than chance. These methods were then adapted for use in squirrel monkeys (*Saimiri sciurieus*; Russell, Early, Painter, & Judge, in prep). Four of five squirrel monkeys tested chose the high-value token significantly more than chance. In all of the aforementioned studies, forgoing the immediately available lower valued food in favor of the high-value token was considered a demonstration of self-control, as it incorporates extra effort (an additional exchange is required) and an inherent time delay (as a result of the additional exchange).

Food exchange tasks are presumably more difficult because they do not utilize tokens as a place-holder for the low-value food. Generally, the use of tokens, or other symbolic characters, in place of food rewards enhances performance (Addessi & Rossi, 2011; Boysen, Mukobi, & Berntson, 1999; Mischel & Moore, 1973), however, this is not always the case (Evans, Beran, Paglieri, & Addessi, 2012). Participants are first given a piece of low-value food, and are required to hold it until they are given an opportunity to exchange it with the experimenter for a more desired reward. But at any time during the trial, they can forfeit the high-value reward by consuming the low-value reward in their possession. For this reason, it is considered a delay-maintenance task. In previous studies, the delay between receiving the low-value food and exchanging it has ranged from seconds to minutes depending on the species tested, the individual, and the size of the reward (e.g., brown capuchins, 10-80 s: Drapier, Chauvin, Dufour, Uhlrich, & Thierry, 2005; Pelé, Micheletta, Uhlrich, Thierry, & Dufour, 2011; Ramseyer, Pelé, Dufour, Chauvin, & Thierry, 2006; long-tailed macaques, *Macaca fascicularis*, 10 s-21 min: Pelé, Dufour, Micheletta, & Thierry, 2010; tonkean macaques, *Macaca tonkeana*, 20 s-2 min

40 s: Pelé et al., 2011; chimpanzees, 15 s–8 min: Beran, Rossettie, & Parrish, 2016; Dufour, Pelé, Sterck, & Thierry, 2007; see Table 1). Many studies required the participants to return the initial food item unadulterated, while some studies allowed them to return a partially consumed reward (indicated in Table 1 by a superscript “b”). Returning the reward untouched likely required more self-control than being able to consume some of it in the delay period. Additionally, when partial consumption was tolerated, participants would be expected to optimize their reward amount by consuming as much of the initial, low-value item as allowed. Therefore, one should be careful when attempting comparisons across these studies, as any interspecific differences that were found should not be taken as a true representation of differing self-control abilities between species, as condoned nibbling could be a confounding factor. With this in mind, results generally show a greater capacity for delay maintenance in Old World species as compared to New World species (see Table 1). However, this is further confounded by a paucity of data for New World monkeys with capuchins being the only species studied (Drapier et al., 2005; Pelé et al., 2011; Ramseyer et al., 2006).

Accumulation task

This delay maintenance paradigm presents participants with a collection of food items that are transferred, one by one, within reach of the participant. The participant may take the food items at any time, however, when the participant does this, the experimenter stops transferring the food items, and the participants only receive the items that had accumulated within their reach during the trial. Thus, delay maintenance, and so, self-control, is measured as the amount of time between when the first item becomes available and when the participant decides to end the trial by taking a food item. The first nonhuman accumulation task procedure was modeled after studies of human children largely conducted by Toner and colleagues between 1977 and 1981 (Toner, 1981; Toner, Lewis, & Gribble, 1979; Toner & Smith, 1977) with one modification; all food that could be acquired during a trial were visible for the entirety of the trial (Beran, 2002). For this task to accurately measure delay maintenance, it is essential that the participant understands that food will continue to accumulate as the trial proceeds, as long as they do not interfere. For this reason, studies typically incorporate some kind of training phase or demonstration trials meant to teach the "rules of the game" to the participant, however, each study accomplishes this in a slightly different way. Demonstration trials are crucial because a failure to understand the premise of the test should not be interpreted as a failure to demonstrate delay maintenance. Apes, Old World monkeys, and New World monkeys have been tested on this task with various degrees of success: chimpanzees, up to 10 min (Beran, 2002; Beran & Evans, 2006; Beran & Evans, 2009; Evans & Beran, 2007a; Evans et al., 2012; Parrish, Perdue, Evans, & Beran, 2013); bonobos (*Pan paniscus*), 50 s max delay interval (Stevens, Rosati, Heilbronner, & Mühlhoff, 2011); orangutans (*Pongo pygmaeus*), 1 min 30 s max delay interval (Beran, 2002; Parrish et al., 2014); rhesus macaques (*Macaca mulatta*), 30 s max delay interval (Evans & Beran, 2007b); long-tailed macaques, *Macaca fascicularis*, 2 min 11 s (Pelé et al., 2010); tonkean macaques, *Macaca tonkeana*, 38 s-1 min 12 s (Pelé et al., 2011); brown capuchins, 2 s-1 min 30 s max delay interval (Pelé et al., 2011; Evans et al., 2012;

Addessi et al., 2013; Beran, Perdue, et al., 2016; Anderson, Kuroshima, & Fujita, 2010); squirrel monkeys, 0-30 s, (Anderson et al., 2010; see Table 2).

In the pioneering nonhuman primate study, Beran (2002) tested four chimpanzees and an orangutan in four differing conditions after an initial training phase. The basic procedure involved moving 20 chocolate pieces from one clear bowl out of the participant's reach to another clear bowl within their reach. The training phase modeled the testing phases in all procedures but decreased the number of chocolate pieces that could be accumulated to 10 and continued until they could wait for all 10 pieces to be transferred (maximum required was four training sessions). The first three phases varied in the length of the delay and in the orientation of the experimenter to the participant. In phase one, the addition of each food piece was delayed by 3 s and the experimenter remained oriented toward the participant. In phase two, the delay was increased to 6 s and the experimenter had to turn and reach behind him to pick up the next piece of food to be transferred. In this way, the experimenter effectively had his back turned to the subject for approximately 3 s of the 6 s delay. For the third phase, the delay was increased to 9 s and the experimenter had to walk away from the participant to collect the next food piece to be transferred, again effectively turning his back to the participant, but this time for a longer duration. The forth phase investigated the effect of food preference on delay maintenance and varied the type of food that accumulated across (not within) trials. The results indicated that the orientation of the experimenter did not influence the apes' ability to demonstrate delay maintenance, nor did differing food preferences. However, any food preference effect could have been obscured by a ceiling effect, since all participants were already close to perfect performance across all test phases.

Evans and Beran (2007b) tested rhesus macaques on a similar task where up to ten grapes could accumulate in their food pan, with a transfer occurring once every 3 s, creating a maximum possible delay of 30 s. There was no mention of training trials for this study, so it seems as though they were able to spontaneously infer that all the food would be transferred, and learned that if they interfered, the accumulation stopped. In a later phase of the study, they found that having a single grape follow nine pieces of low-value food increased the participants' delay tolerance. Though there were marked individual differences, five of the nine participants were able to tolerate the entirety of the delay on several occasions.

Anderson, Kuroshima, & Fujita (2010) modified Evans and Beran (2007) procedure for squirrel monkeys and brown capuchins, however, procedures varied slightly between species. Participants had at least two demonstration trials throughout each testing session, one for the very first trial, and at least one more during the session (capuchins received four, while squirrel monkeys received two). Within blocks of three trials, the duration of the delay between transfer was randomly determined, so that one trial would consist of 1 s delays, another would have 3 s delays, and yet another would have 5 s delays. I believe the randomization of the differing delay lengths was to prohibit the monkeys from developing an expectation of how much time would pass before the next delivery. Evans and Beran (2007) also addressed this concern, but in a different manner. If the participants were not successful in obtaining more than one piece of food across several trials, they moved on to phases that were thought to encourage performance in a stepwise fashion. The first of which presented the participant with a “free” piece of food before the start of the trial. The second included this free piece of food and presented food of increasing sizes, similar to how Evans and Beran (2007) increased the value of the last food item for macaques. One of the four squirrel monkeys demonstrated self-control during the first phase that only included demonstration trials. Another squirrel monkey and two of the four capuchins also did so after the food items increased in size.

Behavioral Inhibition

Given their procedural ease and potential for adaptability to a wide range of species with differing perceptual and motor modalities, behavioral inhibition paradigms are a frequent choice for researchers. Typically, this suite of tasks (including, but not limited to, A-not-B, middle cup, cylinder, swinging door, Plexiglas hole) involve a transparent barrier that the participant must simply reach their hand around or maneuver their body around to retrieve a reward. However, these tasks are contentious within the field. Some researchers use them interchangeably with other self-control tasks, interpreting the results as being representative of the individual’s or species’ overall self-control ability (MacLean et al., 2014). However, others (Beran, 2015) have disagreed with this approach, arguing that behavioral inhibition should not be treated as if it is synonymous with self-control. Both support their opinions with their definition of self-control. Beran (2015) defines self-control as a choice between two options differing in value and/or cost, whereas MacLean et al. (2014) defines self-control as, “the ability to inhibit a prepotent but ultimately counterproductive behavior." Thus, it seems that the nature of the debate may extend beyond the appropriateness of the tasks into the essence of what self-control is (or is not). While there may never be a unanimous consensus as to how self-control should be defined, differences between types of tasks and what they measure can be further investigated.

The A-not-B task examines whether participants can inhibit searching for food in a location where they have found food before when the food is moved to a new location. The participant is first presented with a few familiarization trials, where they watch the experimenter place a reward under the left (or right) most cup of a linear, three cup arrangement. They are then allowed to retrieve the reward. After completing the familiarization trials, the subject receives a single test session in which they watch the experimenter bait the same cup that was baited in the previous trials, however, before the participant has a chance to retrieve the reward, the experimenter conspicuously moves the reward to the other outer-most cup. A correct response is searching for the reward in the new location, thereby demonstrating "self-control" through inhibiting an impulsive, prepotent motor response to search in the old location.

Amici, Aureli, and Call (2008) conducted a battery of tests meant to measure aspects of behavioral inhibition on a wide selection of primates and correlated performance with the differing dietary preferences, social structures, and phylogenies of the difference species. One of their five tasks was the A-not-B task, which they had adapted from Piaget's (1954) experiment with human children. Another was a delay-choice task (intertemporal choice). Performance across the tasks was most closely correlated with social structure, with fission-fusion species demonstrating the highest levels of behavioral inhibition. While the authors never explicitly mention the term "self-control" in the article, they did utilize a traditional self-control task alongside a more controversial one. Additionally, they looked at behavioral inhibition as a measure of adaptive behavioral flexibility, concluding that individuals in fission-fusion societies perform better because they must flexibly demonstrate/inhibit certain behaviors to successfully navigate their ever-changing social surroundings. They go on to say that their results support the social brain hypothesis, which states that species with more complex social lives have evolved more advanced cognitive abilities (Dunbar, 1998). One of the leading theories as to why self-control evolved, closely resembles this interpretation, stating that self-control is one component of the advanced cognitive abilities developed. In fact, as I will discuss shortly, this is the main argument of MacLean et al. (2014).

The cylinder task is similar to the A-not-B task, beginning with a series of familiarization trials where participants watch an experimenter bait a horizontally positioned opaque tube-shaped round cylinder with a food reward and are then allowed to retrieve the reward. These trials function to teach the participant the proper retrieval technique that involves reaching around the front of the cylinder to the side opening. The test trials are procedurally the same, except that the opaque cylinder is replaced with a transparent one. If the subject reaches directly for the reward, instead of maneuvering around the cylinder as they did in the familiarization trials, it is scored as an incorrect response. The cylinder task and similar tasks have been termed detour paradigms because a detour must be taken around the side of the cylinder to obtain the food. Because the participants have presumably learned the correct response through the familiarization trials, incorrect responses are thought to arise from an impulse to reach directly for the reward. Impulsivity is often thought of as the opposite of self-control, and so interpreting this behavioral inhibition as impulsivity, presumably quantifies the lack of self-control in participants.

MacLean et al. (2014) explicitly state that they used the A-not-B and cylinder tasks to measure self-control as a proxy for overall cognitive function, with the ultimate aim of investigating hypotheses regarding the proximate and ultimate causes of the evolution of higher cognitive processes. The results suggest that the proximate cause of self-control is absolute brain volume, whereas the ultimate cause is a variable diet. Not surprisingly, apes were reported to have the highest composite score from both tasks. However, the relative ranks of the other species are intriguing, with apes being followed by domestic dogs, then Old World monkeys, lemurs, rodents, birds, and lastly squirrel monkeys. While these tasks can be considered measures of self-control under the broader definition of self-control as the ability to inhibit ultimately counterproductive behavior, Beran (2015) argues that self-control tasks must consist of three factors in order to for them to actually be testing self-control. These factors are: 1) at least two known choices for participants to choose from, 2) both choices must be valuable, but differentially preferred by the subject, 3) there must be some cost associated with the more preferred reward. Considering these criteria, neither of the tasks utilized by MacLean et al. (2014) would be considered valid tests of self-control. They may meet the first criterion, in that the A-not-B task has three different cups to choose from and it could be argued that the cylinder task incorporates choices of “reach toward” and “reach around.” However, both tasks fail to meet the second criterion because not all of their choices are valuable, because essentially an incorrect response results in no reward. The third criterion is debatable, the A-not-B might not have known options if they are unaware that rewards can be retrieved from cups other than the one used in familiarization trials or if they do not understand the causality of moving the reward. For the cylinder task, previous studies have explicitly stated that the familiarization trials serve to “teach” them the correct response of detouring around the barrier, however, countless familiarization trials would not teach them to reach around a transparent barrier if they do not fully grasp the concept of transparency. Therefore, they would not know that they have two choices from which to choose.

In addition to Beran’s criticisms, a recent study that tested several corvid species on the cylinder task contended that if they had been included in MacLean’s study, the correlations between absolute brain size and self-control would have been nonexistent, as the corvids’ performance on the task rivaled that of great apes despite their smaller brains (Kabadayi, Taylor, von Bayern, & Osvath, 2016). Several studies have since corroborated Kabadayi et al.’s (2016) argument and are thoroughly reviewed by Kabadayi, Bobrowicz, and Osvath (2018) with the general conclusion that specific brain regions or neuronal density/populations are responsible for the expression of behavioral inhibition, not absolute brain size. More importantly, Kabadayi et al. (2018) also review important procedural considerations of the detour paradigm. They support the use of detour tasks, because they are easy to administer and “ecologically relevant,” with animals often needing to detour around things in their natural environment to achieve a desired goal. However, they discuss how many noncognitive factors such as reward visibility, distance from the reward, rearing in an enriched environment, prior exposure to transparent objects, motivation, and age can influence an individual’s performance on detour tasks. Furthermore, they provide evidence that seems to suggest performance inconsistencies on different detour tasks are due to the orientation of the barrier. They briefly mention the inherent issue of transparent objects, stating that participants should have sufficient experience with them prior to testing, however, this topic is better addressed by van Horik et al. (2018). Expanding upon Kabadayi et al. (2018), van Horik et al. (2018) provided evidence that two inhibitory control tasks (cylinder and barrier) were indeed highly influenced by noncognitive factors such as motivation to retrieve the reward and previous experience with transparent objects. However, they go further than Kabadayi et al. (2018), suggesting that the cylinder task may not be an accurate measure of behavioral inhibition, let alone self-control, coining the term “putative inhibitory control tests” to refer to tasks that involve transparent barriers. They suggest that any behavioral inhibition task that utilizes transparent barriers is confounded by the fact that animals would not naturally come into contact with transparent objects in the wild, therefore, there is no reason to expect that they would innately understand their properties. For this reason, any tests that utilize transparent objects could be considered invalid, especially since van Horik et al.’s (2018) results are not the only ones to suggest that exposure to transparent objects (like glass windows used in captive environments) could give certain individuals advantages over others (Bojczyk & Corbetta, 2004; Kabadyi et al., 2016). As suggested by Kabadayi et al. (2018), one way to circumvent this issue, is to give participants ample opportunities to learn the rules of transparent objects before testing begins, however, this has rarely been addressed in experimental designs (e.g., MacLean et al., 2014).

Behavioral inhibition tasks can be advantageous because of their adaptability to a wide variety of species and ease of administration, however, it seems as though one must be extremely wary of noncognitive factors that could affect a participant's performance. The legitimacy of behavioral inhibition tasks as a measure of self-control is still debatable, and arguably, dependent on how self-control is defined in the context of the experiment. Given the claims that behavioral inhibition and other self-control tasks measure equivocal processes, an investigation into the topic is warranted. To the best of my knowledge, no experiments have explicitly addressed this question. Therefore, there is a need for a within-subject design testing individuals on both behavioral inhibition tasks and other, more accepted self-control paradigms, in order to compare performance across tasks.

Aims and Predictions

Except for brown capuchin monkeys, there is a paucity of data regarding the self-control abilities of New World monkeys, and the few studies that do address the topic are difficult to compare with existing paradigms used on Old World monkeys or utilized an experimental paradigm (i.e., intertemporal choice) not included in this review due to its vastly problematic nature (for overview, see Hayden, 2016). Despite a general opinion that squirrel monkeys have relatively short attention spans and are highly distractible (Fragaszy, 1985), some previous experiments indicate that they are viable subjects for tests of self-control abilities (Anderson, Awazu, & Fujita, 2000; Anderson, Awazu, & Fujita, 2004; Anderson et al., 2010; Russell et al., in prep). I tested squirrel monkeys on an array of existing self-control paradigms. These experiments were the first to examine the performance of squirrel monkeys in an exchange task that did not involve the use of tokens. Squirrel monkeys have been previously tested on the other proposed tasks: delay-maintenance (Anderson et al., 2010), A-not-B (MacLean et al., 2014), and cylinder (MacLean et al., 2014). However, I believed there was sufficient reason to warrant replications of the tasks with refined methodologies. Additionally, replication of the tasks allowed for a within-subject design that was used to address the efficacy of the paradigms, and what they might be telling us about self-control and impulsivity.

The delay-maintenance paradigm conducted by Anderson et al. (2010) was designed to explicitly help the participants perform well on the accumulation task. Therefore, the squirrel monkeys were given many affordances that other species did not receive and it was interesting to determine if they were able to learn the task with less prompting. Additionally, the experiment was ended without providing the squirrel monkeys with a task that could be compared to previously tested species. My methodology excluded these training biases in addition to replicating phase one of experiment one from Evans and Beran (2007) to allow for direct comparison across species. The present studies combined aspects of several experiments (Anderson et al., 2010; Beran, 2002; Beran, Perdue, et al., 2016; Evans & Beran, 2007b) to produce a test suitable for squirrel monkeys but more comparable to a broader range of previously completed accumulation studies.

In both the A-not-B and cylinder tasks, squirrel monkeys performed quite poorly when compared with other nonhuman primate species (MacLean et al., 2014). As previously discussed, the field would benefit from experimental evidence explicitly addressing behavioral inhibition tasks in relation to other tests of self-control. Testing squirrel monkeys that have already been tested in a self-control task with varying degrees of success (Russell et al., in prep) allowed me to compare their performance on a more widely accepted self-control paradigm to that of the more contentious behavioral inhibition tasks. If their performance replicated that of the squirrel monkeys in MacLean et al. (2014) it would be the first experiment to directly show the same set of subjects excelling on one self-control task (token exchange/bartering) and performing poorly on the controversial self-control tasks (A-not-B and cylinder). Results would be the first evidence to strongly corroborate Beran's (2015) assertion that these two tasks are not accurate tests of self-control abilities.

Importantly, this was one of the few self-control experiments that incorporated a within-subject design (Addessi et al., 2013; Beran, Perdue, et al., 2016; Parrish et al., 2018; Pelé et al., 2010; Pelé et al., 2011) and the only study that attempted to control for experience effects through counterbalancing tasks within each category. Eisenreich and Hayden (2018) mention that a human delay maintenance task (the marshmallow test) and a delay choice task (intertemporal choice) differentially predict success measures later in life, while Addessi et al. (2013) raised the question of whether delay choice and delay maintenance were comparable tests. As stated earlier, it is thought that delay of gratification is a multifaceted form of self-control, consisting of an initial choice and then fidelity to that choice. They compared the performance of brown capuchins across the two subsets of delay of gratification tasks, delay choice (intertemporal choice) and delay maintenance (accumulation), and found only weak evidence for the tasks being equivalent. Their results partially support the hypothesis that delay choice and delay maintenance are separate components of delay of gratification. Pelé et al. (2011) tested tonkean macaques and brown capuchin monkeys on two different delay maintenance paradigms (food exchange and accumulation), focusing their discussion upon the interspecies comparison rather than performance across tasks within a species. While it was not their focus, their results do show that capuchins were able to wait 2 to 4 times longer in the accumulation task than in the exchange task. However, they did not find similar results for the macaques who waited just as long over both tasks. These results seem to suggest that exchange tasks are more difficult for capuchins, and perhaps other New World monkeys, than accumulation tasks, and thus, the tasks themselves may not be comparable, even though they both address delay maintenance. Therefore, it is possible that these two delay maintenance tasks (food exchange and accumulation) may be not be testing the same aspects of self-control, and further investigation with another New World monkey species could elucidate any differences between these tasks and species performance across them. The experiments conducted explicitly investigated how squirrel monkeys perform across these two delay-maintenance tasks to assess whether they were comparable as well as comparing them to a different set of self-control tasks (behavioral inhibition). Beran, Perdue, et al. (2016) found that training capuchins on one self-control task did not improve performance on another self-control task, however, experience within a specific task did improve performance on that task. While the sample size for these studies was too small to explicitly test for experience effects, counterbalancing the tasks within type (A-not-B with cylinder and food exchange with accumulation) controlled for the possibility that experience within a particular set of tasks allowed subjects to perform better on subsequent tasks of a similar type.

The proposed battery of self-control tasks not only investigated the extent of self-control abilities in squirrel monkeys, but also allowed for the comparison of the paradigms themselves. Addessi et al. (2013) highlights the importance of an investigation of this nature stating that, “a better understanding of what each task exactly measures is necessary to foster our knowledge of origins and mechanisms of delay of gratification.” Two recent reviews on the topic have echoed this sentiment, calling into question various measures of self-control that are often taken to be equivalent (Flessert & Beran, 2018) and bringing attention to the “dearth of validated self-control measures” along with the need to study a wider variety of species (Eisenreich & Hayden, 2018).

In the present studies, the accumulation and food exchange tasks differed between the low- and high-value choices with one mostly incorporating a difference in quantity (accumulation), and the other dealing with a difference in quality (food exchange task), however, one of the four phases of the accumulation task did include a quality distinction as well. While differences in performance across the two tasks used in the present studies (accumulation and food exchange) could not directly be compared in relation to this quality-quantity distinction, it could lend itself to further investigation on what factors may have affected a participant’s motivation to exhibit self-control. Some studies suggest that participants will demonstrate more self-control when rewards differ in quality compared to when they differ quantity (capuchins: Beran, Perdue, et al., 2016; Ramseyer et al., 2006; corvids: Hillemann, Bugnyar, Kotrschal, & Wascher, 2014), suggesting that accumulation may be inherently more difficult than food exchange. The present study allowed for further investigation of this through comparing performance within the accumulation task.

Methods

Subjects and Housing

Six female squirrel monkeys ranging in age from 10 to 25 years participated in the present studies. The squirrel monkeys were socially housed at the Bucknell University Animal Behavior Laboratory in a single group with one other female squirrel monkey who was not be tested due to poor eye sight. A majority of the participants had previous experience using a touchscreen and choosing between baited objects (i.e., cups and tools: Judge, Tomeo, Zander, Powell & Miller, 2015; Marsh, Vining, Levendoski, & Judge, 2015; Painter, Russell, & Judge, in press; Zander, Weiss, & Judge, 2013). It should be noted that transparent cups were used by Zander et al. (2013) and that the monkeys had seen transparent objects before (windows, plastic containers, Plexiglas dividers attached to caging). Five of the six had prior experience with a token exchange paradigm designed to test self-control, in which four successfully demonstrated self-control (Russell et al., in prep).

The captive-bred monkeys were housed in an indoor/outdoor enclosure consisting of an approximately 2.3 x 5.8 x 5.3 (h x w x l) m indoor quarter and an outdoor corncrib measuring 12 m high and 5 m in diameter. The indoor quarter was constructed of plastic paneling and stainless steel welded-wire caging and was subdivided into three approximately equal sized compartments separated by interconnecting doorways. The front of each compartment was constructed of 1 x 2" h x w (2.5 x 5.1 cm) caging wire, which provided an experimenter access to the monkeys for testing. The enclosure contained structures such as swings, shelves, and perches to promote naturalistic locomotion. Enrichment items were continually present. Animals were fed twice daily at approximately 0730 and 1600 hr on a diet of high protein monkey biscuits, fruits, nuts, grains, cereals and vegetables, and water was available *ad libitum*. The research was approved by and complied with Bucknell University IACUC guidelines. The subjects were maintained in accordance to USDA guidelines and husbandry plans were annually submitted to and approved by Bucknell University IACUC.

General Procedures

Testing occurred seven hours after the initial feeding, immediately before the second to control for hunger state which could affect motivation to retrieve the reward. Subjects were separated from their conspecifics in a side compartment of the home enclosure for testing, as in previous experiments. As mentioned, the A-not-B and cylinder tasks were counterbalanced across the subjects to control for possible effects of experience. Delay maintenance and food exchange tasks were also counterbalanced for the same reason. All six subjects were used in the A-not-B, cylinder, and delay-maintenance tasks. The food exchange task was restricted to five subjects because one monkey did not learn to exchange items with the experimenter and was not used in the previously conducted token exchange task. Subjects only received one test session per day, regardless of the task.

Experiments

A-not-B

I replicated Amici et al.'s (2008) procedure, which was a slight modification of Piaget's (1954) experiment with human children, and was used in MacLean et al. (2014) as a test of self-control. The task consisted of a series of familiarization trials immediately followed by ten test trials. For the familiarization trials, three cups, each 3.81 x 4.45 x 5.72 (h x w x l) cm were placed on their side on a platform, 3.81 x 8.89 x 40.64 (h x w x l) cm (Figure 1). A reward (mealworm) was placed in front of one of the outermost cups and then each cup was flipped up sequentially, starting with the baited cup, to cover the reward. The same outermost cup was used across all familiarization trials within subjects, but left and right sides were randomly counterbalanced between subjects. The platform was then pushed forward so the subject could select a single cup. Subjects were required to successfully complete three consecutive trials, by accurately selecting the cup containing the reward, before moving on to the test trials. After the familiarization criterion was met, the subject was given ten test trials immediately afterwards, as a part of the same session. These test trials began exactly like the familiarization trials, but before the platform was pushed forward, the reward was conspicuously moved from the familiar outermost cup to the opposite outermost cup (see Video 1, note the entire platform was painted blue for experimental trials as in Figure 1). The subject's selection was recorded as a success (newly baited cup) or a failure (familiar or middle cup).

Cylinder Task

Many variations of this task have been conducted, however, I replicated MacLean et al.'s (2014) procedure to allow for direct comparison with their results. Subjects were given a series of familiarization trials until a criterion of 4 out of 5 correct responses on consecutive trials was met. For these trials, I visibly baited a horizontal, opaque PVC cylinder that was painted blue (3in long x 1.5in inside diameter) with a mealworm (Figure 2A). The subjects were then given a chance to obtain the reward. A correct response was recorded as a reach to one of the open sides of the cylinder (Figure 3A). Any attempt that involved a reach to the top or front side of the cylinder was recorded as incorrect (Figure 3B). Regardless of the subject's initial response, they were allowed to retrieve the reward. Immediately after criterion was met, ten test trials were performed similarly, except that the opaque cylinder was replaced with a transparent, but otherwise identical PVC cylinder (Figure 2B).

Food Exchange Task

The procedure for this task was based on that of Drapier et al. (2005) and Ramseyer et al. (2006), but was not a complete replication of either. Both studies incorporated the same two familiarization sessions. I replicated the first as a training phase in which the monkeys were given a token and subjected to a five second delay before they could exchange that token for a food reward until they reached a criterion of successfully exchanging in at least nine trials of the 10-trial session. As the participants were already proficient exchangers, this served as a reacclimation phase, as it was approximately two months since they had engaged in any exchange behavior by the start of data collection.

The second familiarization session from Drapier et al. (2005) and Ramseyer et al. (2006) was utilized as my first testing phase, in order to determine if squirrel monkeys would spontaneously exchange food for food with experimenters, whereas most other experiments have incorporated at least one training session to train participants to exchange one extremely low-valued food item for a more desirable one (Drapier et al., 2005; Dufour et al., 2007; Pelé et al., 2010; Pelé et al., 2011; Ramseyer et al., 2006). Low- and high-value rewards for each subject were determined and utilized in a previous experiment (see Table 3 for a summary, Russell et al., in prep). Prior to testing for the current studies, these low- and high-value pairs were verified for all participants, with all still preferring their high-value food significantly more than their low-value food. Monkeys needed to choose the high-value food at least 17 out of 20 trials over two 10-trial sessions to verify their preference (according to a binomial distribution, $p < .01$). On a separate day of testing, it was verified that all participants would still consume 10 pieces of their low-value food when presented alone. Reward values differed qualitatively because evidence from two populations of capuchins as well as corvids suggest that qualitative reward differences are more likely to promote and sustain delay of gratification than quantitative differences (Beran, Perdue, et al., 2016; Hillemann et al., 2014; Ramseyer et al., 2006).

Subjects were shown the low-value food in one hand and the high-value food in the other hand for 3 s, with the food placements semi-randomly counterbalanced between hands across trials, so that the same arrangement did not occur more than twice in a row. Then I gave the monkey the low-value food and immediately held out a cupped hand against the caging, while continuing to show the monkey the high-value food. If the monkey placed the low-value food in my cupped hand, they were given the high-value food. Monkeys were required to return the low-value reward intact, although both studies cited above allowed subjects to consume part of the reward in some experimental phases. Subjects received a minimum of two 10-trial sessions but continued until they met a criterion of 90% successful exchanges over their last two sessions (a criterion similar to that of Ramseyer et al., 2006). Monkeys who reached criterion moved on to the second phase of testing which incorporated time delays before an exchange could be made. The minimum of two 10-trial sessions allows for the comparison of performance between this task and the previous bartering task the participants completed (Russell et al., in prep).

Phase two replicated the first experimental phase of Ramseyer et al. (2006) conducted on brown capuchin monkeys. Participants received 12-trial sessions in which six trials had no delay (exactly like phase one) and six trials of specified increasing time delays (2, 5, 10, 20 s). Delays were measured here, and in subsequent tasks, by counting out the seconds, which was verified with a digital wrist watch. Trial types were semirandomized, so that no more than two trials of the same type (delay vs no delay) were presented sequentially. The purpose of the six no-delay trials was to keep the participants interested in the experiments and facilitate their continued persistence (Ramseyer et al., 2006). During the delay, the high-value food remained visible, however, I did not present a cupped hand for exchange until the specified delay had passed. If the low-value reward was consumed or dropped the trial ended and the time of the failure was recorded. If a participant successfully exchanged in 5 out of 6 delay trials they moved onto the next, longer time delay. A maximum of three 12-trial sessions for each time delay was set. If participants failed to reach the 5/6 criterion in three sessions but exchanged in at least 3 out of 18 total delay trials (16% of trials), they also moved onto the next, longer time delay. Participants that failed to exchange at least 3 times within each specified time delay were not tested further. These criteria were chosen because they were used in Ramseyer et al. (2006). The percentage of successful exchanges was used for analysis, as subjects received a different number of trials/sessions.

Participants who successfully exchanged the low-value food for the high-value food at or above criterion in delay trials were given two, twelve-trial control sessions. Following the same procedure described above, they were given the high-value food and then offered an opportunity to exchange it for the low-valued food. This controlled for the possibility that the participants learned an associative rule such as “always exchange with the experimenter.”

Accumulation Task

Anderson et al. (2010) tested squirrel monkeys on an accumulation task, however, their procedure was created to enhance the monkey's performance, and deviated from existing accumulation task procedures (e.g. Beran, 2002; Evans & Beran, 2007b). My procedure combined aspects of four experiments to remove the majority of the affordances given to the squirrel monkeys that were tested by Anderson et al. (2010) and to allow for comparison to previously tested species.

The apparatus mimicked that of Anderson et al. (2010) and consisted of a rolling tray 42.54 x 57.15 (w x l) cm with two white rectangles, each 4.45 x 22.86 (w x l) cm that were parallel to each other, 17.34 cm apart, but perpendicular to the subject. One rectangle was out-of-reach for the participant and acted as the "store zone" that held all the rewards that could be obtained during a single trial (Figure 4). The rectangle within reach, located 4.61 cm from the front edge of the rolling tray, acted as the "drop zone" to which the rewards from the store zone were transferred. All other food items for subsequent trials were hidden from the participant's view.

Phase one consisted of a maximum of four 4-trial sessions, because this total of 16 trials was comparable to the 15 trials administered to the rhesus macaques by Evans and Beran (2007b). To prepare a trial, 6 pieces of the participant's high-value food (the same used in the food exchange experiment) were placed in the store zone and then the tray was rolled up to the caging. The trial began when the first piece of high-value food was placed into the drop zone. The subsequent food pieces were transferred into the drop zone one at a time, as quickly as possible, so that there was approximately a 1 s delay between pieces. A trial ended when the monkey took a piece of food from the drop zone. At that time, they were allowed to collect all the food that had accumulated in the drop zone while the remaining food in the store zone was removed from view. The empty cart was then rolled away from the caging to begin a 30 s inter-trial interval. This phase was intended to reveal whether or not squirrel monkeys could spontaneously perform on this task similarly to other species tested (Beran, 2002; Evans & Beran, 2007b). All monkeys then moved on to phase two to account for Beran, Perdue, et al. (2016) incorporating demonstration trials at the start of each testing session, allowing the results of the present study to be compared to those of the aforementioned study.

Phase two was procedurally similar to phase one except for the inclusion of demonstration trials and a variation in delay length. This was a direct replication of phase one of Anderson et al. (2010) and consisted of ten 8-trial sessions. Trials 1 and 5 were demonstration trials, where the rolling cart was pulled away from the caging while the trial proceeded. All of the food pieces were moved from the store zone to the drop zone with a 3 s delay between transfers before the cart was pushed forward to the caging, allowing the monkeys to retrieve the rewards. Pulling the cart away from the caging prevented them from ending the trial early because they could not reach the drop zone, effectively demonstrating that all food pieces would eventually be moved within their reach. As previously described, trials 2-4 and 6-8 had one each of 1, 3, and 5s delays randomly ordered, to prevent the participants from forming expectations. After completing 10 sessions, all monkeys moved on to phase three.

Phase three was procedurally the same as phase two above, with the inclusion of the demonstration trials, except that the last piece of food in the sequence was valued higher that the preceding five pieces. More specifically, the first five pieces of food were low-value and the sixth piece of food was of high-value, utilizing the low- and high-valued food for each participant from the food exchange task. Results from Evans and Beran (2007b) suggested that the qualitative discrepancy in reward values would motivate them to allow the trial to finish before reaching for food. Criterion for this phase was accumulating all food pieces available for accumulation in three consecutive sessions. Monkeys that performed at criterion on this phase moved on to phase four, which was identical to phase two of testing, to determine if they could continue to delay gratification when all of the rewards were highly valued. The "free" pieces of food that were given to the participants at the start of each trial in phase two of Anderson et al. (2010) were not incorporated into this study because it was not shown to effect performance.

The fifth and final testing phase increased the number of items that could be accumulated during a trial to 10, with a 3 s delay between transfers. Therefore, this final testing phase was a direct replication of Evans and Beran's (2007b) experiment with rhesus macaques, and allowed for comparison between most of the species previously tested.

The number of food items obtained were recorded. Participants were compared as to how many pieces of food they were able to accumulate as a percentage of the total food pieces they could have accumulated. The percentage was used because participants received a different number of trials. However, the first piece of food for each trial was not included in analysis since this piece signified the start of a trial and all subjects would, by default, accumulate at least that one piece.

Data Analyses. Chi-Square tests were used to compare the frequency of successful monkeys on the cylinder and A-not-B tasks to those of MacLean et al. (2014). Wilcoxon signed rank tests were used to compare performance across phases in both the accumulation task and the food exchange task. Performance across tasks was compared by visually inspecting the ranks each animal received for each task. Ranks were assigned based on number of correct responses (A-not-B, cylinder), how quickly they met criterion (token exchange, food exchange – no delay phase), and the percentage of food pieces exchanged (food exchange – overall) or accumulated (accumulation).

Results - Behavioral Inhibition Paradigms

A-not-B and Cylinder Tasks. Ten test trials were administered, however, for the comparisons made to previous studies that only conducted one test trial (Amici et al., 2008; MacLean et al., 2014) only the first test trial was considered. Individual and species performance in the first test trial was summarized for all subjects in Table 4, along with a comparison to the squirrel monkeys' performance from MacLean et al. (2014). Analyses were conducted identically to MacLean et al. (2014). Performance was recorded as correct (C) or incorrect (I), with correct receiving a score of 100 and incorrect receiving 0. Composite scores were calculated by taking the average of the A-not-B and cylinder scores for each individual. Individual scores were averaged to provide a group score for each task. All participants chose correctly on their first A-not-B test trial resulting in an average group score of 100 ($N = 6$), which is sharply contrasted by the average species score from the 19 squirrel monkeys in MacLean et al. (2014) at a low of 15.8. A 2 x 2 Chi Square test examining the relationship between population (Bucknell vs MacLean) and number of successful monkeys (Yes or No) was statistically significant, $X^2(1, N = 25) = 14.04, p < .001$, indicating Bucknell monkeys were successful more than expected and MacLean moneys were successful less than expected. Four individuals chose correctly on their first cylinder test trial, for an average group score of 66.66, which was noticeably higher than the 33.37 species average score from MacLean et al. (2014), however, a 2 x 2 Chi Square test was not significant, $X^2(1, N = 25) = 1.65, p = .21$, indicating no population differences. One of the two individuals (Debi) that was unsuccessful on her first trial took considerably longer to meet the criterion to pass the familiarization trials than the rest of the participants. The composite score for the present study was 83.33 and differed markedly from the 24.7 from MacLean et al. (2014). The composite results also indicated population level differences, 2 x 2 Chi Square test $X^2(1, N = 25) = 6.18, p < .03$.

Individual and species performance for all ten test trials was summarized for all subjects in Table 5. Individual scores were calculated by averaging the scores they received for each trial of each particular task. Each individual score then contributed to the “average species score” reported. The comparison to MacLean et al. (2014) could not be made here due to the differing number of test trials. One difference was that two individuals chose incorrectly on their second and sixth trials in the cylinder task, depressing the average cylinder group score to 96.67. Another notable point was that one individual continued to perform poorly throughout the cylinder testing session, only making a correct response on the second and tenth trials. Perhaps unsurprisingly, this individual was the one who took much longer to meet the familiarization criterion than all others (Debi). Despite her performance the average group score of 85 was still higher when considering all test trials compared to 66.66 (first trials only), and the group composite score of 90.83 was higher still.

Discussion - Behavioral Inhibition Paradigms

A-not-B and Cylinder Task. Based on MacLean et al.'s (2014) results, I predicted that squirrel monkeys would not perform well on both the A-not-B and cylinder tasks, despite their successful performance on the bartering paradigm (Russell et al., in prep). If this was the case, it would have been the first experimental evidence to show that behavioral inhibition tasks were not equivalent to other, more accepted tests of self-control. Unfortunately, the present studies did not provide conclusive evidence to contribute to the Beran vs. MacLean debate as to whether or not performance on these tasks would correlate to performance on other, more widely accepted self-control tasks. Contrary to my predictions based off of MacLean et al. (2014), all but one subject performed well, including the individual who had failed to show self-control in the preliminary token exchange task. When attempting to assign ranks for the two behavioral inhibition tasks, there was not enough individual variation in performance, even when considering all ten testing trials, to break the four-way tie for first place, precluding the analysis necessary to assess performance across tasks. Considering the degree of difference between the results from the present study compared to that of MacLean et al. (2014) one might be tempted to attribute the higher performance in this study to experience with self-control tasks. However, the disparity becomes even more surprising considering the MacLean et al. (2014) species scores included three individuals that had past experience on more traditional self-control tasks (e.g., Anderson et al. 2010), on which two successfully demonstrated self-control. However, MacLean et al. (2014) did not provide individual data, precluding further investigation as to how those specific individuals performed on the tasks, nor was their familiarity, or lack thereof, with transparent objects addressed. Furthermore, their study included individuals from two different populations of squirrel monkeys and the authors noted that the average scores of the populations did not differ significantly, which was unusual since the present study which was essentially a third population performed significantly better. These facts, coupled with minimal explicit data regarding the squirrel monkeys previous experience with cognitive tasks, made it difficult to hypothesize what could be responsible for the disparity in scores between the two populations tested in MacLean et al. (2014) and the one tested here.

The present data do support the arguments made in Kabadayi et al. (2018) and van Horik et al. (2018) in that many cognitive and noncognitive factors can affect performance on behavioral inhibition tasks in general, and more specifically, the cylinder task. For this reason alone, it would be wise for researchers to avoid these tasks when investigating anything other than behavioral inhibition or behavioral flexibility, despite their convenience and adaptability. Additionally, one could argue that the A-not-B task is further confounded by a memory component in that the subject must remember where they last saw the reward. While subjects must only remember that information for a matter of seconds, it should still be considered when assessing the validity of the task, as memory capabilities vary between species. I also believe this task was intentionally misleading, in that the familiarization trials essentially train the subjects to look for food in that particular cup and they had never received food from any other cup previously. It is possible that looking for the reward in the familiar cup results from a habituated foraging strategy or risk avoidance rather than true behavioral inhibition. From this perspective, the task resembles a reversal-learning task more so than one of self-control. It would be interesting to see how the lower performing species would perform on this task if the containers were transparent, which would control for the issue of memory and habituation. One would assume that they would just go for the visible reward, but if they do not, it would clearly indicate a fatal flaw in the task. Transparency would not be an issue here, as it was in the cylinder task, because the familiarization trials with the transparent cups should provide the subject with enough experience with transparent objects. A test of this nature could help address the validity of the task and might provide parameters as to when it should be used, if it should be used at all. Future studies could also investigate the validity of these tasks by determining if the specific brain regions implicated in behavioral inhibition tasks are also activated in more accepted self-control tasks. Kabadayi et al. (2016) suggested that specific brain regions and neuronal densities/populations are responsible for successful performance in these tasks, rather than absolute brain volume (as suggested by MacLean et al. 2014).

Results - Food Exchange Task

All subjects performed at criterion on the training phase in just one session. Criterion was successfully exchanging at least nine of ten high-value tokens for the high-value food after a 5 s delay. As a reminder, this training phase is comparable to the first set of familiarization sessions in Draiper et al. (2005) and Ramseyer et al. (2006).

All subjects also performed at criterion in the first testing phase (exchanging the low-value food for the high-value food with no delay) within three to ten sessions (*M* = 7.6 sessions, *SD* = 2.8, Table 6), with criterion being the successful exchange of at least nine of ten low-value food pieces two sessions in a row. Individuals' percent exchanges for these no delay sessions ranged from 10 – 86.67 % (*M* = 58.99%, *SD* = 31.08, Table 6), and all subjects moved on to the next phase incorporating 2 s delays. Ranks, from 1 – 5, were assigned for this phase separately, to allow for a direct comparison to the previously conducted token exchange task (Table 7). The following analyses include only the 6 delay trials per session (the 6 no delay trials per session were omitted from analysis, as they only functioned to keep participants engaged and performance was nearly 100%). As a reminder, the criterion for advancement here (borrowed from Ramseyer et al., 2006), was exchanging in at least five of the six trials of a single session, or exchanging in at least three trials over the 18 trials (6 trials per session, over three sessions) of a given time delay. At the 2 s delay, four of the five subjects had a perfect performance (100%

exchange) in a single session and immediately progressed from the 2 s delay to the 5 s delay. The fifth subject met criterion, exchanging in 5 of the 6 exchange trials, after only two sessions with a 75% exchange rate (group: *M* = 95%, *SD* = 11.18). All subjects met criterion in the 5 s delay phase within one to three sessions, ranging from 66.67 – 83.33% exchange (*N* = 5, *M* = 69.44, *SD* = 18.42). Performance in the 7 s delay phases decreased slightly, although all subjects met criterion within two to three sessions. Two met the 5 out of 6 criterion and three met the 3 out of 18 criterion, with percent exchange ranging from 16.67 – 75% (*N* = 5, *M* = 49.44, *SD* = 24.09). In the 10 s phase, one subject failed to move on to the next delay (15 s), exchanging only twice over the 18 trials. Of the remaining four subjects, two met the 5 out of 6 criterion and two met the 3 out of 18 criterion, allowing them to be tested at the 15 s time delay. Percent exchange ranged from 11.11 – 83.33% (10 s phase, *N* = 5, *M* = 44.44, *SD* = 30.17). Only one of the four individuals that were tested at the 15 s delay length met the minimum criterion to move on to the next delay length (20 s), exchanging in 3 of the 18 trials (16.67%). Another individual exchanged only twice (11.11%) and the other two subjects did not exchange at all (15 s phase, group analysis, *N* = 4, *M* = 6.95, *SD* = 8.33). Overall, this is a significant decrease in performance from the 10 s phase, *Z* = -2.023, *N* = 5, *p* = .043. The one individual that progressed to the 20 s delay length exchanged twice (11.11%), failing to meet the 3 out of 18 criterion necessary to test at a delay length of 25 s. Therefore, the overall delay that was tolerated by the squirrel monkeys ranged from 10 – 20 s (Table 6). Four of the five monkeys performed above chance on the control sessions, with three individuals never exchanging the high-value food for the low-value food, and one individual exchanging only twice ($p < 0.001$, binomial test, needed to have 6 or fewer exchanges out of 24 to be significant), the fifth performed at chance on the control test exchanging the high-value food with the experimenter 17 times ($p = .585$, binomial test, needed to have nine or fewer exchanges out of 30 to be statistically significant), however, it should be mentioned that she was cycling at the time and displaying atypical behavior such as refusing to participate in the trials and discarding any value food she was given during the testing time for three days in a row, including the food she had exchanged for. Given the occurrence of this rare behavior, I suspect that her performance on the controls was a fluke and I did not remove any of her data from the food exchange analyses. Animals were ranked 1-5, with one being the highest, based on who was exchanging most often at the longest time delay (Table 7).

Discussion - Food Exchange Task

I predicted that the individuals who excelled in the previously conducted token exchange (bartering) paradigm would also excel in the food exchange task, because they had already demonstrated self-control under a similar situation, therefore the results from that experiment were considered alongside those of the present studies. Anecdotally, some subjects who obtained the low-value food during bartering trials, attempted to barter with the low-value reward (i.e., they tried to place the low-value food into my hand by thrusting their arm through the caging while holding the low-value food), providing further reason to suspect that they would be successful in the food exchange task.

The present experiment was the first to show that squirrel monkeys were capable of exchanging less preferred food with an experimenter to receive a more preferred food. As seen in Table 7, the ranks assigned for the initial testing phase (No Delay) did not match the ranks assigned for overall performance on the task (including delay trials of various lengths). Ranks were, however, identical to those from the token exchange task, which had a minimal (no) delay period. The result was slightly surprising, as the food exchange task was presumably more difficult because the participant must actively inhibit themselves from eating an immediately available food reward in their possession as opposed to choosing an immediately available food reward that was not yet in their grasp. Yet, we see that performance was comparable in trials of equal length delays – in this case no delay. I am curious to see if this pattern would hold across differing delay lengths, however, a token exchange experiment incorporating delays has not yet been conducted in squirrel monkeys. Further evidence from the present study actually suggests that the food exchange task was less difficult than the token exchange task, contradicting previous studies (Addesi & Rossi, 2011; Boysen, Mukobi, & Berntson, 1999; Mischel & Moore, 1973) because one subject (Arlene) exchanged tokens only minimally (in 2 of 50 trials) but exchanged her low-value food for the high-value food in 36 of 90 no delay trials and continued exchanging as delay length increased to 15 s. I hypothesize that this may be due to the prevalence of the high-value reward at the time of the decision to either maintain self-control or to defect. In other words, in the token exchange task, the subject only sees the low-value food and the high-value token, as the high-value food was located behind the experimenter while the monkey made her choice. Conversely, in the food exchange task, the high-value reward was held in front of the subject for the entirety of the delay length, perhaps increasing the motivation to maintain self-control. An alternate explanation for increased performance in the food exchange task compared to the token exchange task was that the food exchange task removed some of the cognitive demands on the participant. The participant would not have to remember the association between the high-value token and the high-value food at the time of the decision. A third explanation was that when faced with a choice between a consumable item and an unconsumable item, the response of reaching toward an available piece of food was more salient. The present study did not provide enough evidence to support one of these hypotheses over the other, however, I plan to conduct a follow up experiment to determine if Arlene's performance on the token exchange task will be higher after her success on the food exchange task. If her performance does indeed increase, it might suggest that a cognitive factor initially hindered her performance and has now been learned. No change in performance would not provide us with any new information as to why there was increased performance on the food exchange task compared to the token exchange task.

As mentioned earlier, performance on the token exchange task did not predict performance as to how long of a delay an individual would tolerate before defecting and consuming the low-value reward. This was best exemplified by Violet's versus Echo's performance. Violet was ranked number one in both the token exchange and the no delay food exchange phase, reaching criterion in just 3 sessions. However, her performance sharply decreased, from 66.67% to 0%, between the 10 s and 15 s delay lengths in the food exchange task, which might suggest that she simply decided it was no longer advantageous to her to maintain the delay. Echo took longer to reach criterion in the no delay food exchange phase, meeting criterion in seven sessions, however, Echo's rate of performance declined more gradually, with some exchanges still occurring at the 20 s delay. I suspect that this could be explained by a difference in ability versus motivation. For example, the ability to learn that maintaining a delay was "more" profitable was most likely separate from a participants' propensity or willingness to actually tolerate that delay. Another factor that could have been at play here, is that the value of the reward varied subjectively for each individual participant. As the time delay increased, the value of the reward decreased at differing rates for each individual, a concept that has been addressed in multiple publications (e.g., Hayden, 2016) and in Beran's (2018) current operational definition of self-control. It was unlikely that this was a result of more experience (i.e., additional trials for Echo) because those trials did not incorporate a delay and therefore could not have contributed to any shaping of a longer delay maintenance.

Violet was not the only subject whose performance sharply decreased from the 10 s to 15 s delay length, in fact, all subjects performances significantly decreased. It was possible that a more gradual increase in delay length would have protected against this drop in performance, however, I was concerned that continuing to increase delays by such small increments would shape the subjects' performance. In other words, one could have made the argument that subjects were being trained to delay even though the intention was to measure their inherent self-control abilities.

Results indicated that squirrel monkeys have self-control levels that were comparable to some capuchin groups (Table 1). For example, the capuchin population in Ramseyer et al. (2006) maintained self-control for a range of 10 – 20 s when they were not allowed to nibble the reward before exchanging. Considering that these capuchins had much more experience with delayed food exchanges throughout the experiments conducted in Ramseyer et al. (2006) and were also participants in a previous food exchange task (Drapier et al., 2005), it was quite impressive that naïve squirrel monkeys performed just as well as capuchins did, especially with evidence from Beran, Perdue, et al. (2016) that indicates experience on a particular self-control task leads to increased performance on that task in the future. In fact, this effect could be seen when comparing the studies denoted by a superscript "a" in Table 1, indicating that the same population of capuchins were used in each study, with obvious increases in performance from one study to the next. Additionally, some of those individuals would not delay their gratification for more than 10 s, meaning that two of the squirrel monkeys in the present study (Arlene and Echo) actually performed better than some capuchins, despite having less experience. I suspect that if the squirrel monkeys were given more experience, their performance would be equal to that of capuchins. As discussed earlier, apes and Old World monkeys had much longer delay times than squirrel monkeys and capuchins (Table 1), but at the very least, the results of the current study indicate that squirrel monkeys are valid subjects for tests of self-control and should not be discounted based on their apparent "distractibility" and "short attention spans" (Fragaszy, 1985). It would be interesting to test whether squirrel monkeys perform successfully on this task if the rewards differed quantitatively by exchanging a small piece for a large piece, as in chimpanzees (Beran et al., 2016; Dufour et al., 2007), macaques (Pelé et al., 2010; Pelé et al., 2011), and capuchins (Drapier et al., 2005; Ramseyer et al., 2006; Pelé et al., 2011).

Results - Accumulation Task

All subjects were given phases one and two, however, after phase two, one subject (Debi) was dropped from the experiment because she was having dexterity issues that precluded her from removing the reward from the store zone even though she was trying to do so. For this reason, her performance was not included in the analyses for this task.

In phase one, where subjects were given 16 trials without any demonstrations, none were successful in accumulating any food pieces other than the one that marked the start of the trial. For subsequent phases, the percentage of food pieces accumulated out of the total number of pieces available without counting the initial piece to start the trial was calculated for each phase broken down by each time delay length (1, 3, and 5 s). Ranks were assigned for each delay length and they did not differ markedly, nor were there any interesting anomalies worth noting. Therefore, I collapsed the data across time delay lengths. The percentage of food items accumulated, meaning the number of food items accumulated out of the total possible food items that could have been accumulated (5 pieces, since the first marked the start of the trial and was not included in analysis) were reported for each phase. Performance in phase two, where two demonstration trials were included in each session, was extremely low, with subjects accumulating only 0.33 – 7%

of the available rewards (Table 8), despite the inclusion of the demonstration trials in which the cart was pulled away from the caging to allow all possible food items to accumulate without the subject ending the trial early. Demonstration trials served to show the subjects what would happen if they did not interfere with the accumulation of food, and were not included in analyses for any phase. In general, subjects that succeed in accumulating some food items in this phase did not begin doing so until session seven, indicating that it took multiple sessions for them to learn the task contingencies.

Phase three was noteworthy because the rewards differed by value with the first five pieces were low value and the sixth and final piece was of high value. Performance immediately increased during this phase for all subjects, $Z = -2.03$, $N = 5$, $p < .05$ (Table 8). In fact, three of the five subjects met criterion, which was three trials in a row with perfect performance, which allowed them to progress to phase four rather quickly. Violet and Gwen had accumulation percentages of 100% and 97.78%, respectively, over their first three sessions, Echo had 96% over her first seven sessions and Arlene and Cora completed all ten sessions with 77.33% and 85.67%, respectively.

At the start of phase four, which was essentially a replication of phase two with demonstration trials and all high-valued rewards, individual performances decreased for all subjects, with percentages ranging from 0 – 71.33%. The result was a significant decrease in the group's performance from phase three to phase four, $Z = -2.02$, $N = 5$, $p < .05$. However, individual performances increased from phase two to phase four, for all but one individual, Violet, who accumulated two items in phase two, and no items in phase four. This change in performance trended toward significance, $Z = -1.75$, $N = 5$, $p =$

.08, and if the sample size was larger it may have reached statistical significance. The increase in performance indicated that either the additional experience or alternative reward contingency, or perhaps both, improved their ability to demonstrate self-control in the accumulation task (Table 8).

For this reason, each subject was assigned two separate ranks for the accumulation task, one for overall performance and one for performance on phase three alone, this was to determine if varying the reward differences qualitatively rather than quantitatively affected performance on this task. As seen in Table 7, this difference played a substantial role in performance as every subject's rank was reassigned, the most surprising being that the highest ranked individual became the lowest ranked, and vise-versa.

Due to Violet' s lack of delaying gratification at all during all 10 sessions (60 trials) of phase four, I did not test her in phase five. Phase five consisted of two sessions, each with two trials, where subjects could accumulate up to 10 pieces of high-valued food, with a delay of 3 s between each delivery. The test was conducted to make a direct comparison with Evans & Beran (2007b). Again, Arlene stood out, as she was the only subject whose performance increased from phase four to five, 71.33 to 86.11%, where all others decreased at least some amount (Table 8).

Discussion - Accumulation Task

Based on Anderson et al.'s (2010) results, I suspected that the squirrel monkeys would be capable of succeeding in the accumulation task, to some extent. Meaning that at least some of the participants should progress to the final phase of testing. The data allowed for the direct comparison across Old World and New World monkey species as well as apes, and could lend itself to the discussion of how and why self-control abilities evolved. Furthermore, the within-subject design allowed me to speculate what each paradigm was testing, if the paradigms were testing similar aspects of self-control, and how we might expect other species' performance to vary across tasks. For example, Pelé et al.'s (2011) results suggested that squirrel monkeys might be able to tolerate longer delays during the accumulation task than in the food exchange task.

In phase one, the squirrel monkeys did not spontaneously delay gratification long enough to accumulate any reward pieces. While the result answers a question raised by Anderson et al. (2010), and allows for comparison between Evans & Beran (2007b), it could simply be that there was not enough information for the squirrel monkeys to learn the "rules of the game," and their self-control abilities should not be diminished for not understanding the task at hand. The inclusion of demonstration trials (phase two) and reward variances (phase three) allowed them to conceptualize the rules of the task, which was shown through increased performances in phases four and five. However, it was unclear if the demonstration trials would have been enough on their own to have led to the marked increase between phases two and four with more sessions, or if seeing the reward disparity in a different context (phase three) afforded them the opportunity to conceptualize the rules of the game in a manner that could later be generalized to the previous condition. In other words, the present experiment cannot determine if continuing phase two test sessions for additional experience would have eventually led to the level of performance seen in phase four, or if having experience with low- and high-value rewards was critical in their understanding of the task. The marked increase in performance from phase two to phase four in the current study, was not displayed by rhesus macaques (Evans & Beran, 2007b). Perhaps the macaques had already fully learned the task, and the differing phase contingencies only revealed differences in motivation to delay gratification, whereas in the present study results likely indicated that learning took place during phase three. There was no means to determine if the learning occurred because of, or in conjunction with, the motivation increase.

The change in performance from phase two (all high-valued food) to phase three (five low-valued foods followed by one high-value) parallels that of rhesus macaques in experiment one, phase two of Evans & Beran (2007b). It also corroborates previous findings that suggest animals are more likely to delay gratification when rewards differ qualitatively (low/high value) rather than when they differ quantitatively (small/large: Beran, Perdue, et al.,2016; Ramseyer et al., 2006; Hillemann et al., 2014). Future work should endeavor to uncover the reason for this pattern, as it was possible that the qualitative difference was more salient, or simply more desirable, which increased motivation to delay. For example, if a child was given the chance to delay gratification by forgoing accumulating crackers for a cupcake, I would suspect higher performance than if he/she were given the chance to accumulate multiple cupcakes. More importantly, it showed that the type of reward disparity was a critical factor in predicting a subject’s performance across tasks of self-control. When comparing the rankings assigned in phase three of the accumulation task to that of the token and food exchange tasks with no delay

(Table 7), both of which also addressed qualitative differences in rewards, there was an almost perfect consistency of ranks across the tasks.

Overall, squirrel monkeys performed quite well on the accumulation task, surpassing the squirrel monkeys and even the capuchins of Anderson et al. (2010) with more subjects delaying gratification regularly. It was possible that squirrel monkey performance in this study was better due to the qualitative reward difference used in phase three, because Anderson et al. (2010) only differed rewards quantitatively. It was also revealed that the majority of training biases provided by Anderson et al. (2010) were necessary for the present population of squirrel monkeys to delay gratification on the accumulation task. Performance in the present study was comparable to some capuchin populations (Addessi et al., 2013; Evans et al., 2012) but fell short of others (Pelé et al., 2011; Beran, Perdue, et al. 2016), which was impressive considering that they, again, had less experience than the subjects of these studies. Perhaps even more surprising, Arlene's performance in phase five was comparable to rhesus macaques (Evans & Beran, 2007b), although it did not compare to the performance of other Old World monkey species that successfully maintained delays of over a minute (long-tailed macaques, Pelé et al., 2010; tonkean macaques, Pelé et al., 2011; chimpanzees, Beran et al., 2016; Dufour et al., 2007; see Table 1).

General Discussion

While the debate over the differing definitions of self-control is one of a philosophical nature that cannot necessarily be answered through empirical experiments, the present studies have shed some light on what these paradigms are testing and how they should be interpreted. As discussed earlier, behavioral inhibition tasks should be treated with caution, as it is likely that they are not accurate measures of self-control, supporting the assertions in Beran's (2015) opinion article.

The squirrel monkeys in the present study clearly could maintain delays much longer (1.5 to 3 times as long) in phase three of the accumulation task than in the food exchange task which was similar to the capuchins in Pelé et al. (2011), who waited 2 to 4 times longer. These results suggest that the tasks are not of comparable natures and should not be thought of as interchangeable. However, it should be noted that macaques performed similarly on both (Pelé et al., 2011) and future studies should investigate possible reasons for this difference in performance. I suspect that two factors were contributing to the difference in tolerated delay lengths, either separately, or in conjunction. The first was that it was more challenging to have to physically hold on to the low-value food rather than just inhibiting the impulsive reaction to reach for it. The second was that being able to watch the rewards accumulate had some effect on the subject's motivation to delay gratification, perhaps being able to watch each piece be transferred helped to bridge the delay. A future study could address this possibility with a modification to the accumulation procedure used here and create a task that more closely resembles food exchange. The experimenter could start the trial off with a piece of the low value reward in the drop zone with the high value reward in the store zone for durations equal to those used in the food exchange task. If performance resembled that of the accumulation (longer delays tolerated), it would suggest that it was more challenging for a subject to have to physically hold the food. If performance resembled that of the food exchange task, it would support the hypothesis that seeing the food accumulate played a role in their performance. A third, but unlikely reason, was that the subjects were sensitive to the total amount of food they received. However, as discussed, individuals tended to delay longer for qualitatively different rewards as opposed to quantitatively different, so this scenario should not be an exception. If the accumulation task was considered in full, not just phase three, as discussed immediately above, the results supported my prediction that the accumulation would be inherently harder than food exchange in the current studies due to the nature of the difference of their rewards, which was discussed in the accumulation section.

I believe that both of these tasks, accumulation and food exchange, are valid tests of self-control although their results are not directly comparable to each other, especially if the qualitative/quantitative nature of the rewards differ. However, future studies could help determine whether we see these differences between task performance because the tasks elicit different levels of motivation, in addition to addressing which of these should be utilized by experimenters moving forward.

It was thought that the nature of the test, whether it be delay choice or delay maintenance, measured fundamentally different aspects of self-control, with one measuring an animal’s propensity to initially delay gratification and the other measuring their ability to maintain that choice once it was made. Addessi et al. (2013) supported this hypothesis, showing only weak evidence for task equivalency between an intertemporal choice task (delay choice) and an accumulation task (delay maintenance). However, the results of the current studies do not support that hypothesis. In fact, at first there did not appear to be any correlation within a subject's performance across tasks overall, especially in relation to delay maintenance tasks (denoted by "DM" in Table 7) and delay choice tasks ("DC"). But, when I started to separate phases in consideration of other factors, patterns started to emerge. What appeared was that the qualitative/quantitative nature of the rewards played an important role in predicting the subjects' performance on the tasks, at least when the delay was quite short (Table 7). As delay length was increased (as seen in the food exchange, overall column) the nature of the reward was no longer a predictive factor. Future studies could address how performance on hybrid delay tasks, briefly mentioned in the introduction of this paper, might compare to performance on delay maintenance and delay choice tasks, keeping in mind the importance of the nature (quantity/quality) of the choice.

Anecdotally, results support those of Beran, Perdue, et al. (2016) in that experience on one specific self-control task did not necessarily improve performance on subsequent self-control tasks of varying types. The highest ranked monkey in each task (Table 7) was one of the subjects that started the experiment with that task as opposed to the other. More specifically, Arlene, Gwen, and Violet started with the accumulation task and moved on to the food exchange task after they had completed every phase of the accumulation task. As Table 7 shows, Arlene performed the best on the accumulation task overall, and Violet performed best on phase three (which was notable due to its qualitative nature, as discussed above). Echo and Cora started with the food exchange task before receiving the accumulation task and Echo received the highest overall rank for the food exchange task. Violet was the only subject to improve her rank from her original task (overall accumulation) to her second task (food exchange “No Delay” and “Overall”) but based on her performance on phase three of the accumulation task, I believe this improvement occurred due to the change from quantitative reward differences to those of qualitative differences. She tended to perform extremely well when she was working to optimize a grape reward over a carrot, but only performed minimally, if at all, when she was working to optimize the number of grape pieces she received. Her performance in the accumulation task showed that she was capable of delaying gratification, and therefore understood the reward contingencies. The result begs the question of how she made the decision of when it was “worth it” for her to delay and when it was not, as witnessed in the sharp changes in her performance in phases two through four of the accumulation task, going from basically 0% to 100% and back to 0%.

Finally, the present studies added to the existing literature pool of self-control studies, specifically addressing the paucity of data for New World monkey species in a way that allowed for the results to be directly compared to previous studies. The results of the present study support that the presence of marked individual differences in self-control abilities is ubiquitous throughout the literature pool in both humans and animals. Unsurprisingly, apes perform better than Old World monkeys, who in turn perform better than New World monkeys. However, it is worth mentioning again how well the squirrel monkeys in the present study performed despite their lack of experience in self-control tasks, making them promising subjects for future studies of self-control. Furthermore, the presence of self-control abilities in apes, Old World monkeys, and New World monkeys suggests that the capacity evolved before the divergence of ancestral New World monkeys 40 million years ago (Goodman, 1999).

References

Addessi, E., & Rossi, S. (2011). Tokens improve capuchin performance in the reverse–reward contingency task. *Proceedings of the Royal Society of London B: Biological Sciences*, *278*(1707), 849-854.

Addessi, E., Paglieri, F., Beran, M. J., Evans, T. A., Macchitella, L., De Petrillo, F., & Focaroli, V. (2013). Delay choice versus delay maintenance: Different measures of delayed gratification in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *127*(4), 392-398.

Anderson, J. R., Awazu, S., & Fujita, K. (2000). Can squirrel monkeys (*Saimiri sciureus*) learn self-control? A study using food array selection tests and reverse-reward contingency. *Journal of Experimental Psychology: Animal Behavior Processes*, *26*(1), 87-97.

Anderson, J. R., Awazu, S., & Fujita, K. (2004). Squirrel monkeys (*Saimiri sciureus*) choose smaller food arrays: Long-term retention, choice with nonpreferred food, and transposition. *Journal of Comparative Psychology*, *118*(1), 58-64.

Anderson, J. R., Kuroshima, H., & Fujita, K. (2010). Delay of gratification in capuchin monkeys (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, *124*(2), 205-210.

Amici, F., Aureli, F., & Call, J. (2008). Fission-fusion dynamics, behavioral flexibility, and inhibitory control in primates. *Current Biology*, *18*(18), 1415-1419.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*(6), 351–355.

Beran, M. J. (2002). Maintenance of self-imposed delay of gratification by four chimpanzees (*Pan troglodytes*) and an orangutan (*Pongo pygmaeus*). *The Journal of General Psychology*, *129*(1), 49–66.

Beran, M. J. (2015). The comparative science of "self-control": What are we talking about?. *Frontiers in Psychology*, *6*, 51. doi:10.3389/fpsyg.2015.00051

Beran, M. J. (2018). *Self-control in animals and people* (1st ed.). Academic Press. Cambridge, MA.

Beran, M. J., & Evans, T. A. (2006). Maintenance of delay of gratification by four chimpanzees (*Pan troglodytes*): The effects of delayed reward visibility, experimenter presence, and extended delay intervals. *Behavioural Processes*, *73*(3), 315-324.

Beran, M. J., & Evans, T. A. (2009). Delay of gratification by chimpanzees (*Pan troglodytes*) in working and waiting situations. *Behavioural Processes*, *80*(2), 177-181.

Beran, M. J., & Evans, T. A. (2012). Language-trained chimpanzees (*Pan troglodytes*) delay gratification by choosing token exchange over immediate reward consumption. *American Journal of Primatology*, *74*(9), 864-870.

Beran, M. J., Perdue, B. M., Rossettie, M. S., James, B. T., Whitham, W., Walker, B., ... & Parrish, A. E. (2016). Self-control assessments of capuchin monkeys with the rotating tray task and the accumulation task. *Behavioural Processes*, *129*, 68-79.

Beran, M. J., Rossettie, M. S., & Parrish, A. E. (2016). Trading up: Chimpanzees (*Pan troglodytes*) show self-control through their exchange behavior. *Animal Cognition*, *19*(1), 109-121.

Bojczyk, K. E., & Corbetta, D. (2004). Object retrieval in the 1st year of life: Learning effects of task exposure and box transparency. *Developmental Psychology*, *40*(1), 54-66.

Boysen, S. T., Mukobi, K. L., & Berntson, G. G. (1999). Overcoming response bias using symbolic representations of number by chimpanzees (*Pan troglodytes*). *Animal Learning & Behavior*, *27*(2), 229-235.

De Petrillo, F., Micucci, A., Gori, E., Truppa, V., Ariely, D., & Addessi, E. (2015). Self-control depletion in tufted capuchin monkeys (*Sapajus spp.*): Does delay of gratification rely on a limited resource? *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01193

Drapier, M., Chauvin, C., Dufour, V., Uhlrich, P., & Thierry, B. (2005). Food-exchange with humans in brown capuchin monkeys. *Primates*, *46*(4), 241-248.

Dufour, V., Pelé, M., Sterck, E. H. M., & Thierry, B. (2007). Chimpanzee (*Pan troglodytes*) anticipation of food return: Coping with waiting time in an exchange task. *Journal of Comparative Psychology*, *121*(2), 145-155.

Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *6*(5), 178-190.

Eisenreich, B. R., & Hayden, B. Y. (2018). Cognitive science: Persistent apes are intelligent apes. *Current Biology*, *28*(4), R160–R162. https://doi.org/10.1016/j.cub.2018.01.020

Evans, T. A., & Beran, M. J. (2007a). Chimpanzees use self-distraction to cope with impulsivity. *Biology Letters*, *3*(6), 599-602.

Evans, T. A., & Beran, M. J. (2007b). Delay of gratification and delay maintenance by rhesus macaques (*Macaca mulatta*). *The Journal of General Psychology*, *134*(2), 199-216.

Evans, T. A., Beran, M. J., Paglieri, F., & Addessi, E. (2012). Delaying gratification for food and tokens in capuchin monkeys (*Cebus apella*) and chimpanzees (*Pan troglodytes*): When quantity is salient, symbolic stimuli do not improve performance. *Animal Cognition*, *15*(4), 539-548.

Fisher, K. A. (2014). Brown capuchins (*Cebus apella*) exhibit self-control on a delay of gratification food exchange task when food options differ qualitatively. (Unpublished honors thesis). Bucknell University, Lewisburg, Pennsylvania, USA.

Flessert M., & Beran M.J. (2018) Delayed Gratification. In Vonk J., Shackelford T. (eds) Encyclopedia of Animal Cognition and Behavior. Springer, Cham. https://doi.org/10.1007/978-3-319-47829-6_1489-1

Fragaszy, D. M. (1985). Cognition in squirrel monkeys: A contemporary perspective. In *Handbook of squirrel monkey research* (pp. 55-98). C. L. Coe & L. A. Rosenblum (Eds.), Springer, Boston, MA.

Goodman M. (1999). The genomic record of Humankind's evolutionary roots. *American Journal of Human Genetics*, *64*(1), 31–39. doi:10.1086/302218

Hackenberg, T. D. (2009). Token reinforcement: A review and analysis. *Journal of the Experimental Analysis of Behavior*, *91*(2), 257-286.

Hayden, B. Y. (2016). Time discounting and time preference in animals: A critical review. *Psychonomic Bulletin & Review*, *23*(1), 39-53.

Hillemann, F., Bugnyar, T., Kotrschal, K., & Wascher, C. A. (2014). Waiting for better, not for more: Corvids respond to quality in two delay maintenance tasks. *Animal Behaviour*, *90*, 1-10.

Judge, P. G., & Essler, J. L. (2013). Capuchin monkeys exercise self-control by choosing token exchange over an immediate reward. *International Journal of Comparative Psychology*, *26*(4), 256-266.

Judge, P. G., Tomeo, O., Zander, S. L., Powell, B. & Miller, L. E. (2015). Squirrel monkeys demonstrate the identity concept on a match-to-sample task. *American Journal of Primatology*, *77* (S1), 86.

Kabadayi, C., Bobrowicz, K., & Osvath, M. (2018). The detour paradigm in animal cognition. *Animal Cognition*, 21(1), 21-35.

Kabadayi, C., Taylor, L. A., von Bayern, A. M. P., & Osvath, M. (2016). Ravens, New Caledonian crows and jackdaws parallel great apes in motor self-regulation despite smaller brains. *Royal Society Open Science*, *3*(4). https://doi.org/10.1098/rsos.160104

MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... & Boogert, N. J. (2014). The evolution of self-control. *Proceedings of the National Academy of Sciences*, *111*(20), E2140-E2148.

Marsh, H. L., Vining, A. Q., Levendoski, E. K., & Judge, P. G. (2015). Inference by exclusion in lion-tailed macaques (*Macaca silenu*s), a hamadryas baboon (*Papio hamadryas*), capuchins (*Sapajus apella*), and squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, *129*(3), 256-267.

Mischel, W., & Moore, B. (1973). Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology*, *28*(2), 172-179.

Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7, pp. 249– 292). New York, NY: Academic Press.

Painter, M. C., Russell, R. C., & Judge, P. G. (2019). Capuchins (*Sapajus apella*) and squirrel monkeys (*Saimiri sciureus*) fail to attend to the functional spatial relationship between a tool and a reward. *Journal of Comparative Psychology*.

Parrish, A. E., Emerson, I. D., Rossettie, M. S., & Beran, M. J. (2016). Testing the glucose hypothesis among capuchin monkeys: Does glucose boost self-control?. *Behavioral Sciences*, *6*(3), 16. doi:10.3390/bs6030016

Parrish, A. E., James, B. T., Rossettie, M. S., Smith, T. R., Otalora-Garcia, A., & Beran, M. J. (2018). Investigating the depletion effect: Self-control does not waiver in capuchin monkeys. *Animal Behavior and Cognition*, *5*(1), 118–138.

Parrish, A. E., Perdue, B. M., Evans, T. A., & Beran, M. J. (2013). Chimpanzees (*Pan troglodytes*) transfer tokens repeatedly with a partner to accumulate rewards in a self-control task. *Animal Cognition*, *16*(4), 627-636.

Parrish, A. E., Perdue, B. M., Stromberg, E. E., Bania, A. E., Evans, T. A., & Beran, M. J. (2014). Delay of gratification by orangutans (*Pongo pygmaeus*) in the accumulation task. *Journal of Comparative Psychology*, *128*(2), 209-214.

Pelé, M., Dufour, V., Micheletta, J., & Thierry, B. (2010). Long-tailed macaques display unexpected waiting abilities in exchange tasks. *Animal Cognition*, *13*(2), 263-271.

Pelé, M., Micheletta, J., Uhlrich, P., Thierry, B., & Dufour, V. (2011). Delay maintenance in Tonkean macaques (*Macaca tonkeana*) and brown capuchin monkeys (*Cebus apella*). *International Journal of Primatology*, *32*(1), 149-166.

Piaget, J. (1954). *The construction of reality in the child.* Oxford, UK: Basic Books.

Rachlin, H., & Green, L. (1972). Commitment, choice and self-control. *Journal of the Experimental Analysis of Behavior*, *17*(1), 15–22.

Russell, R., Early, K., Painter, M., & Judge, P. (In prep). Squirrel monkeys (*Saimiri sciureus*) exhibit self-control during a token exchange task.

Ramseyer, A., Pelé, M., Dufour, V., Chauvin, C., & Thierry, B. (2006). Accepting loss: The temporal limits of reciprocity in brown capuchin monkeys. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1583), 179-184.

Stevens, J. R., Rosati, A. G., Heilbronner, S. R., & Mühlhoff, N. (2011). Waiting for grapes: Expectancy and delayed gratification in bonobos. *International Journal of Comparative Psychology*, *24*(1), 99-111.

Toner, I. J. (1981). Role involvement and delay maintenance behavior in preschool children. *The Journal of Genetic Psychology*, *138*(2), 245-251.

Toner, I. J., Lewis, B. C., & Gribble, C. M. (1979). Evaluative verbalization and delay maintenance behavior in children. *Journal of Experimental Child Psychology*, *28*(2), 205-210.

Toner, I. J., & Smith, R. A. (1977). Age and overt verbalization in delay-maintenance behavior in children. *Journal of Experimental Child Psychology*, *24*(1), 123-128.

van Horik, J. O., Langley, E. J., Whiteside, M. A., Laker, P. R., Beardsworth, C. E., & Madden, J. R. (2018). Do detour tasks provide accurate assays of inhibitory control? *Proceedings of the Royal Society B*, *285*(1875), 20180150.

Zander, S. L., Weiss, D. J., & Judge, P. G. (2013). The interface between morphology and action planning: A comparison of two species of New World monkeys. *Animal Behaviour*, *86*(6), 1251-1258.

Table 1.
Summary of previous food exchange tasks

Species	Variable(s) Considered	Range of Delay Tolerated	Reference
Chimpanzees	Quantitative exchanges (rewards 2, 4, 8, 40x size). Delay length.	15 s - 8 min	Dufour et al., 2007
Chimpanzees	Reward type, size, and location. Delay length.	most subjects tolerated the max. 5 min delay	Beran et al., 2016
Long-tailed macaques	Quantitative exchanges (rewards 2, 4, 8, 40x size). Delay length.	10 s – 21 min, 20 s	Pelé et al., 2010
Tonkean macaques	Quantitative exchanges (rewards 2, 4, 8, 40x size). Delay length.	20 s - 2 min 40 s	Pelé et al., 2011
Brown capuchins[a]	Qualitative/quantitative exchanges. Multiple exchanges in a row.	most subjects tolerated 5 s delay[b]	Drapier et al., 2005
Brown capuchins[a]	Qualitative/quantitative exchanges. Multiple exchanges in a row. Delay length.	10 - 20 s 640 s[b]	Ramseyer et al., 2006
Brown capuchins[a]	Quantitative exchanges (rewards 2, 4, 8, 40x size). Delay length.	10 - 80 s	Pelé et al., 2011
Brown capuchins	Gradually increased delay times (1 s increments)	6 - at least 35 s	Fisher, 2014
Squirrel monkeys	Gradually increased delay times (2, 3, or 5 s increments)	10 -2 0 s	Present Study

Note. [a] = same population of subjects used in experiments. [b] = partial consumption of low-value food before exchange allowed.

Table 2.

Summary of previous accumulation tasks

Species	Variable(s) Considered	Range of Delay Tolerated	Reference
Chimpanzees and an orangutan	Delay length (3, 6, 9 s). Orientation of experimenter to the subject (facing, turning back, walking away from). Qualitative Rewards.	All subjects maxed out the delay period (3 min)	Beran, 2002
Chimpanzees	Presence of a distractor (toy)	3 s - 18 min (max delay)	Evans & Beran, 2007a
Chimpanzees	Reward visibility. Experimenter visibility.	0 s - 2.5 min (max delay)	Beran & Evans, 2006
Chimpanzees	Whether they had to "work" for the reward or not	50 s - 7 min, 43 s (max delay)	Beran & Evans, 2009
Chimpanzees	Food or tokens accumulating	6 s - 2 min (max delay)	Evans et al., 2012
Chimpanzees	Identity of social partner	All subjects maxed out the delay period	Parrish et al., 2013
Bonobos	"Reliability" of experimenter	2 of 4 subjects maxed out the delay period (50 s)	Stevens et al., 2011
Orangutans	Delay length. Reward visibility.	10 s - 1 min 30 s (some maxed out the delays in phase one)	Parrish et al., 2014
Rhesus macaques	Effect of quantity, quality, and number of items	3 s - 30 s (max delay)	Evans & Beran 2007b
Long-tailed macaques		30 s - 2 min 11 s	Pelé et al., 2010
Tonkean macaques	Reward size. Delay length.	38 - 72 s	Pelé et al., 2011
Brown capuchins	Reward size. Delay length.	33 - 42 s	Pelé et al., 2011

Brown capuchins	Food or tokens accumulating. Number of items available	2 - 28 s	Evans et al., 2012
Brown capuchins	Food or tokens accumulating. Number of items available	2 - 28 s	Adddessi et al., 2013
Brown capuchins	Number of items were transferred at a time	0 - 1 min 30 s (max delay)	Beran, Perdue, et al., 2016
Brown capuchins	Delay length.	30 s (max delay)	Anderson et al., 2010
Squirrel Monkeys	Delay length.	30 s (max delay)	Anderson et al., 2010
Squirrel Monkeys	Qualitative/Quantitative reward effects. Delay length.	30 s (max delay)	Present Study

Table 3.

Subjects' high- and low-value food pairings and past experience with self-control tasks

Subject	Low-value food	High-value food	Successful in token exchange paradigm?
Arlene	Corn Kernel	Mealworm	No
Cora	Carrot	Mealworm	Yes
Echo	Carrot	Grape	Yes
Gwen	Cheerio	Mealworm	Yes
Debi	Cheerio	Grape	N/A
Violet	Carrot	Grape	Yes

Table 4.

Performance on first test trial of the behavioral inhibition paradigms

Subject	A-not-B	Cylinder	Composite Score
Arlene	C	C	100
Cora	C	C	100
Debi*	C	I	50
Echo	C	C	100
Gwen	C	C	100
Violet	C	I	50
Average Score for Species (Present), First test trial only	100	66.66	83.33
Average Scores for Species (MacLean et al., 2014)	15.8	33.37	24.7

Note. Composite scores were calculated by averaging the "scores" from the A-not-B and Cylinder tasks (100 is awarded for correct and 0 for incorrect performance). C = Correct. I = Incorrect. * = subject who took considerably longer to reach criterion in the familiarization trials on the cylinder task.

Table 5.

Performance on all ten test trials of the behavioral inhibition paradigms

Subject	A-not-B	Cylinder	Composite Score
Arlene	100	100	100
Cora	100	100	100
Debi*	90	20	55
Echo	90	100	95
Gwen	100	100	100
Violet	100	90	95
Average Scores for Species (Present), All Test Trials	96.67	85.00	90.83

Note. Composite scores were calculated by averaging the "scores" from the A-not-B and Cylinder tasks (100 is awarded for correct and 0 for incorrect performance). * = subject who took considerably longer to reach criterion in the familiarization trials on the cylinder task.

Table 6.
Summary of food exchange task

	Training Phase	No Delay	% Return	2 s Delay		% Return	5 s Delay			% Return	7 s Delay			% Return
Subject	# correct	Sessions to crit.		# of ex. (max 6)			# of ex. (max 6)				# of ex. (max 6)			
Arlene	10	9	10.00	6		100	3	6		75.00	4	5		75.00
Cora*	9	10	47.00	6		100	5			83.33	0	0	3	16.67
Echo*	10	7	75.75	4	5	75	4	3	5	66.67	2	2	6	55.56
Gwen	10	9	75.55	6		100	4	6		83.33	2	2	2	33.33
Violet	10	3	86.67	6		100	1	2	4	38.89	4	4	4	66.67
Mean		7.6	58.99			95				69.44				49.44
SD		2.8	31.08			11.18				18.42				24.09

Note. Criterion for moving on to the next time delay was exchanging in at least 5 out of 6 trials in a single session or exchanging in at least 3 trials out of the 18 possible trials for each time delay. Controls are out of 12 unless otherwise specified. * = subject was tested on this task before being tested on the accumulation task. # of ex. (max 6) = Number of exchanges made in each 6-trial session, with each session represented by one column.

Table 6 Continued.

	10 s Delay			% Return	15 s Delay			% Return	20 s Delay			% Return	Controls		
Subject	# of ex. (max 6)				# of ex. (max 6)				# of ex. (max 6)				S 1	S 2	(redo)
Arlene	3	3	1	38.89	0	1	1	11.11					1	1	
Cora*	2	2	0	22.22	0	0	0	0					0	0	
Echo*	5			83.33	1	2	0	16.67	0	1	1	11.11	0	0	
Gwen	1	1	0	11.11									3	5 (of 6)	9
Violet	3	5		66.67	0	0	0	0					0	0	
Mean				44.44				6.95							
SD				30.17				8.33							

Table 7.

Summary of ranks across tasks

Task	Token Exchange	Food Exchange		Accumulation	
Subcategory	No Delay	No Delay	Overall	Phase 3	Overall
Task Type	DC	DM	DM	DM	DM
Reward Disparity	Quality	Quality	Quality	Quality	Mostly Quantity
Arlene	5	5	2	5	1
Cora	4	4	4	4	3
Echo	2	2	1	3	2
Gwen	3	3	5	2	4
Violet	1	1	3	1	5

Note. DC = Delay Choice. DM = Delay Maintenance.

Table 8.

Summary of accumulation results: total percentage of rewards accumulated per phase

Subject	Phase 2	Phase 3	Phase 4	Phase 5
Arlene*	3.67	77.33	71.33	86.11
Cora	3	85.67	42.33	30.56
Echo	7	96.19	44.33	41.67
Gwen*	0.33	97.78	20.67	13.89
Violet *	0.67	100	0	n/a

Note. * = subject was tested on this task before being tested on the food exchange task.

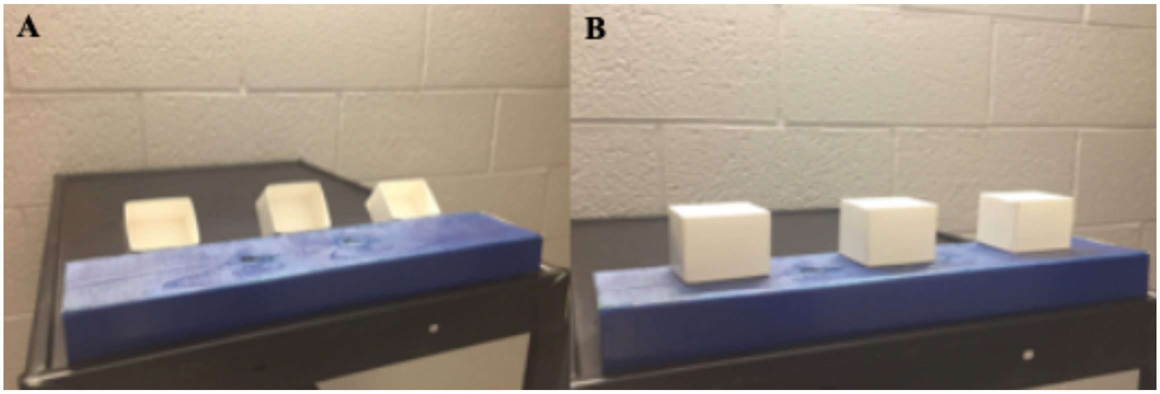

Figure 1. A-not-B apparatus with cups on their side (A). For familiarization trials a reward was placed under an outermost cup and then all cups were flipped up (B). Refer to Video 1 for a demonstration of the test trial.

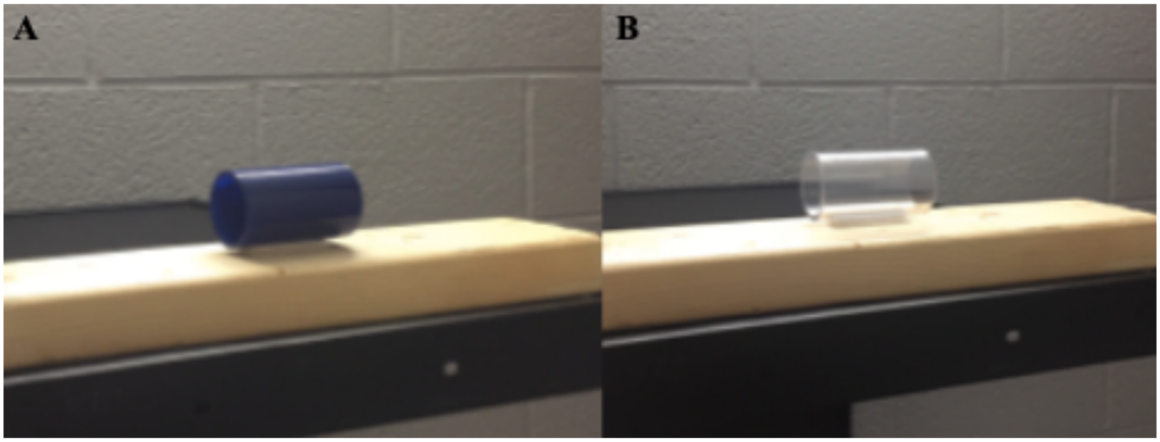

Figure 2. Cylinder task apparatus. The opaque cylinder used in familiarization trials (A) and the transparent cylinder used in test trials (B).

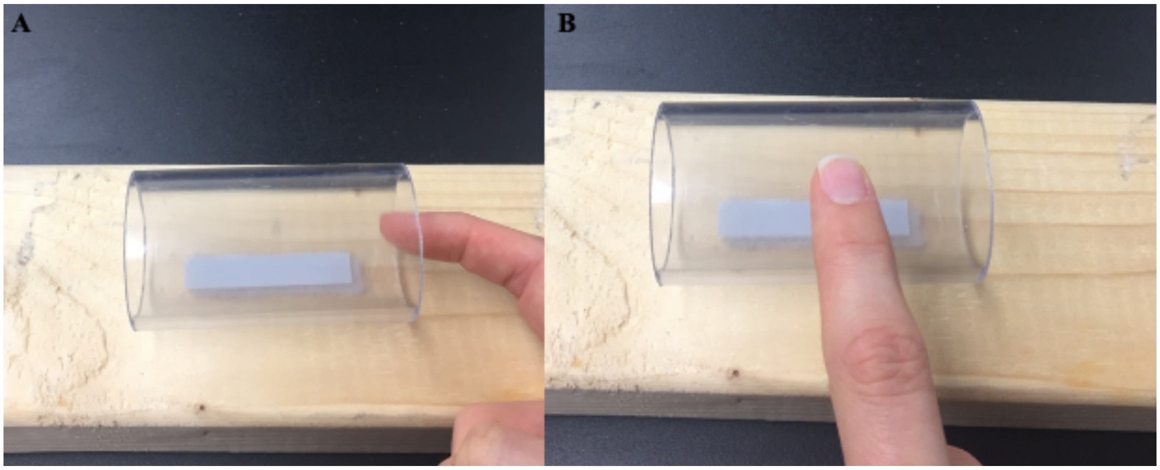

Figure 3. Experimenter demonstrating a correct response (A) and an incorrect response (B).

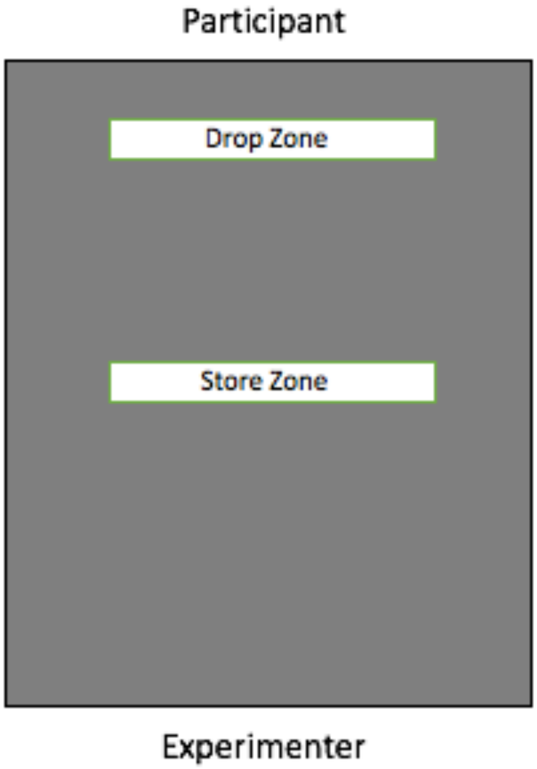

Figure 4. Diagram of accumulation apparatus. Gray rectangle denotes the top of the rolling tray. “Drop zone” is within the participants reach, the “store zone” is not.

Video 1. Demonstration of a test trial for the A-not-B task.