ORIGINAL PAPER

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

T. A. Evans · M. J. Beran · F. Paglieri · E. Addessi

Abstract Capuchin monkeys have been tested for the capacity to delay gratification for accumulating rewards in recent studies and have exhibited variable results. Meanwhile, chimpanzees have consistently excelled at this task. However, neither species have ever been tested at accumulating symbolic tokens instead of food items, even though previous reports indicate that tokens sometimes facilitate performance in other self-control tasks. Thus, in the present study, we tested capuchin monkeys and chimpanzees for their capacity to delay gratification in a delay maintenance task, in which an experimenter presented items, one at a time, to within reach of an animal for as long as the animal refrained from taking them. In Experiment 1, we assessed how long capuchin monkeys could accumulate items in the delay maintenance task when items were food rewards or tokens exchangeable for food rewards. Monkeys accumulated more food rewards than they did tokens. In Experiment 2, we tested capuchin monkeys and chimpanzees in a similar accumulation test. Whereas capuchins again accumulated more food than tokens, all chimpanzees but one showed no difference in performance in the two conditions. These findings provide additional evidence that chimpanzees exhibit greater self-control capacity in this task than do capuchin monkeys and indicate that symbolic stimuli fail to facilitate delay maintenance when they do not abstract away from the quantitative dimension of the task. This is consistent with previous findings on the effects of symbols on self-control and illuminates what makes accumulation a particularly challenging task.

Introduction

Self-control requires foregoing some immediate favorable outcome for the sake of obtaining a more valuable outcome later. This is a useful capacity for humans, and it is essential to planning, appropriate social interactions, physical health and well-being (Baumeister and Vohs 2004; Mischel 1974). Comparative research suggests that the capacity for self-control is not uniquely human and that there is continuity across species with regard to both self-control and its opposite, impulsivity (Logue 1988). In the present study, we examine self-control in two non-human primate species and address different issues that may lead to variability in performance such as methodological variations and species differences.

In the most common forms of self-control testing (called the inter-temporal choice, temporal discounting, or delay choice task), an organism is faced with one of two choices, to be made in the present: Take a small (or less preferred) reward now or wait for a large (or more preferred) reward that will be available later. Once the choice is made, it cannot be altered at any time during the trial. A large variety of organisms have been tested in this delay choice paradigm, including insects (e.g., Cheng et al. 2002), birds (e.g., Chelonis et al. 1994; Logue and Pena-Correal 1985; Mazur 2007), rodents (e.g., Green and Estle 2003; Tobin et al. 1993; van Haaren et al. 1988), non-human primates (e.g., Addessi et al. 2011; Stevens et al. 2005; Tobin et al. 1996),

T. A. Evans (✉) · M. J. Beran
Language Research Center,
Georgia State University, Atlanta, GA, USA
e-mail: theodore.evans@gmail.com

F. Paglieri · E. Addessi
CNR, Istituto di Scienze e Tecnologie della Cognizione,
Rome, Italy

and humans (e.g., Green et al. 1994; Lawyer et al. 2010; Logue et al. 1996).

In another type of self-control test (called the delay of gratification or delay maintenance task), an organism must maintain a course of action in the face of continual competition from the available, impulsive response. In this test situation, even after an initial choice to delay gratification, the immediate reward is continuously accessible throughout the trial, and only inhibition of impulsive responses to the immediate reward leads (eventually) to the delayed reward. Although this may be a less common assessment of self-control capacity (especially in the non-human literature), the delay maintenance task is more relevant to naturally occurring self-control situations in which there are multiple opportunities to exhibit impulsivity. An example for humans would be saving money in an interest-bearing account. Withdrawals can be made at any time, but to obtain the greatest reward, such responses must be delayed. For non-humans, consider the example of the subordinate male that must inhibit the urge to court a preferred female while in the presence of a dominant male. To avoid an aggressive encounter, the subordinate male must continually refrain from consorting with the female until the dominant male leaves.

There is evidence that non-human animals can exhibit delay maintenance in the laboratory, although performance varies across species. Tests of delay maintenance initially developed for children (e.g., Mischel et al. 1989; Toner and Smith 1977) have been modified for use with animals (e.g., Beran et al. 1999; Grosch and Neuringer 1981; Killeen et al. 1981). In one of these tests (hereafter referred to as the accumulation task), subjects are presented with a series of valuable items, one-by-one, as long as they refrain from taking the increasingly larger reward pile. Once they take the accumulated reward, no further items are presented. Thus, every time an experimenter adds a new item to the reward pile, the subject has the choice between taking what has been presented (and ending the trial), or continuing to wait, potentially maximizing the amount of reward to be obtained.

Only a few animal species have been tested using this paradigm, and large differences have been found between those species that have been tested. Great apes have been tested most often and have shown the most promise. Chimpanzees (*Pan troglodytes*) and an orangutan (*Pongo pygmaeus*) have accumulated food items for several minutes before succumbing to the temptation to eat the items (Beran 2002; Beran and Evans 2006, 2009), and chimpanzees have exhibited strategic responses such as self-distraction to facilitate longer delay (Evans and Beran 2007a). Also, bonobos (*Pan paniscus*) have demonstrated a capacity to wait up to 50 s to accumulate 10 food items, but only two of four tested subjects managed to perform reliably in this task

with an inter-reward interval of 5 s (Stevens et al. 2011). Non-ape species have been less successful in this task. African gray parrots (*Psittacus erithacus*) have exhibited limited capacity to delay gratification for even 2–5 s in this paradigm (e.g., Vick et al. 2010). Also, other primate species have not been able to match the performance of apes in this task (squirrel monkeys, *Saimiri sciureus*, Anderson et al. 2010; rhesus macaques, *Macaca mulatta*, Evans and Beran 2007b; but see Pelé et al. 2011 for more promising performance in Tonkean macaques, *Macaca tonkeana*).

Second only to apes, the most work with this paradigm has been done with capuchin monkeys (*Cebus apella*). In one study with this species, two of four monkeys eventually learned to delay gratification for approximately 10–20 s (on average), but they required three different training procedures to do so (Anderson et al. 2010). In another study, three of seven monkeys succeeded in an initial training phase and, in the test phase, accumulated items for approximately 30–40 s, on average (Pelé et al. 2011). Most recently, a group of 14 capuchin monkeys was tested without a formal training phase and these monkeys underperformed in comparison to both of the previous studies (Mean = 2.71 s; data from Addessi et al., submitted). Methodological differences across these studies are a possible explanation of the different levels of performance exhibited by capuchin monkey groups in the accumulation task, as all of these tasks used different delays between item deliveries, different item numbers, and different training procedures. However, another likely explanation for these differences is within species variation, such as individual differences in inherent behavioral tendencies and/or differences in life/ experimental history. This amount of variation within species may be within the normal range considering that there was at least a little overlap in performance between these capuchin monkey studies, but no overlap in performance between the capuchin monkey studies and the ape studies. Regardless of the explanation, this variability of results calls for additional studies on capuchin monkeys with the accumulation task, like the one presented here.

Another question with regard to delay maintenance behavior is whether it is influenced by characteristics of the rewards/outcomes. Presumably, delaying gratification for rewards with directly perceivable prepotent qualities (e.g., visible foods) would be more difficult than delaying gratification for rewards or outcomes that are conveyed more abstractly (e.g., represented by text or images). Previous research has investigated the possibility that substituting symbols or tokens for food items in self-control tests can lead to greater performance; however, these studies have mainly been conducted with a different test paradigm than delay maintenance (but see Beran et al. 1999). For example, consider chimpanzees' performance in the reverse-reward contingency task in which a smaller or less preferred food

option must be selected in order to obtain a larger or more preferred food option (Boysen and Bernston 1995). In this task, the chimpanzees almost always respond to the larger set of food items (and therefore receive the smaller amount of reward), presumably because the larger set inherently has a greater appeal. However, when the sets of food items are replaced by cards displaying Arabic numerals that represent the amount of available food items, chimpanzees exhibit much greater success in responding to the choice option of lower value (Boysen et al. 1999). Here, the symbols have no inherent appetitive qualities, so it is much easier for chimpanzees to not point to the option of larger value. Another study revealed that substituting less abstract stimuli (e.g., rocks) for food rewards in the choice options does not have the same facilitation effect of performance in this task, presumably because it fails to hide the tempting feature, that is, quantity, of the suboptimal option (Boysen 2006).

Additional studies have been conducted in this paradigm that have aided in determining what features substitute-items must possess in order to facilitate individuals' performance. For instance, Vlamings et al. (2006) tested all four great ape species with sets of food items hidden beneath different colored opaque containers and found that this concealment of the food items at choice time resulted in greater selection of the smaller option. Addessi and Rossi (2011) tested capuchin monkeys in the reverse-reward task with different types of tokens representing sets of food items. In one task variation, individual food items were each replaced by "low-symbolic" token objects so that the choice options consisted of two different quantities of token objects. In a second variation, whole choice options were each represented by one "high-symbolic" token that was associated with a particular quantity of food items. This study replicated the findings of Boysen et al. (1999; Boysen 2006; Boysen and Bernston 1995) in that only the high-symbolic tokens, which concealed from direct perception both the appetitive qualities and physical amounts of the rewards, facilitated reversed-reward performance in comparison to the same test employing sets of visible food items.

In the present study, we addressed two main questions. First, we wanted to test whether substituting tokens for food items in the accumulation task would influence capuchin monkeys' performance, given evidence of such an effect in previous studies using the reverse-reward paradigm. In Experiment 1, we presented capuchins different versions of the accumulation task in which the test items were either food items (Phases 1 and 2) or tokens (Phase 3 and 4): This was the first time tokens were used in an accumulation task with this species, and more generally with non-human animals. The nature of the accumulation task requires the use of low-symbolic tokens, that is, objects that have a one-to-one correspondence with food items and can,

thereby, be accumulated in a similar fashion. Using this type of token removes the immediate presence of food but retains the saliency of reward quantity, thus allowing one to investigate what exactly makes accumulation a challenging task for animals. If capuchins' performance is impaired by the physical presence of food rewards within their reach, performance should improve when accumulating non-edible tokens. If, on the contrary, animals have trouble inhibiting a grasping response toward the growing pile of rewards, low-symbolic tokens should not facilitate performance in this case, similarly to what happens in the reverse-reward contingency task.

Our second question was how these capuchin monkeys and a group of chimpanzees would perform in equivalent accumulation tasks. Given the performance differences previously found between different colonies of the same monkey species on this task, it is reasonable to expect considerable differences in performance between different primate species, and it is interesting to check whether these differences mirror those observed in delay choice tasks (for a discussion, see Addessi et al. 2011). Chimpanzees previously have been tested in different versions of the accumulation task, but always with some kind of preferred food as the type of accumulating item. Although chimpanzees succeeded in these tasks, it is unknown how they would behave in a similar task involving accumulating tokens. Therefore, in Experiment 2, we conducted a test very similar to the one used in Experiment 1, but with both capuchin monkeys and chimpanzees. Testing the capuchins in a second experiment provided the advantage of allowing the capuchins to acquire experience with the task before being evaluated alongside the previously tested chimpanzees. This also allowed us to assess the validity of the capuchins' performance in Experiment 1, particularly by examining the effect of food and token task presentation order on monkeys' accumulation performance. We hypothesized that capuchin monkeys would exhibit slightly greater accumulation performance in comparison to Experiment 1, given their added experience with the task. We also hypothesized that chimpanzees would perform at least as well as in previous experiments involving similar tasks. Finally, as in Experiment 1, we hypothesized that tokens would have the same effect (or lack thereof) on delay of gratification in both species, independent of each species' general delay of gratification capacity.

Materials and methods

Subjects

In Experiments 1 and 2, we tested eight capuchin monkeys housed at the Language Research Center (LRC, Georgia

State University, Atlanta, USA): Drella (age 20 male), Gabe (age 12 male), Griffin (age 13 male), Liam (age 7 male), Lily (age 13 female), Logan (age 5 male), Nala (age 8 female), and Wren (age 8 female). Monkeys were group housed but were separated into individual $33 \times 46 \times 61$ cm enclosures for testing. The test area included four side-by-side test enclosures (each separated from the next by a distance of 66 cm) and so monkeys maintained visual and auditory access to other monkeys even while physically separated for testing. Monkeys had 24-h access to water and were fed manufactured chow following morning and afternoon test sessions, as well as various fruits and vegetables between 1,600 and 1,800 h. All monkeys had a large amount of experience with manual and computerized testing in several different cognitive and psychomotor tasks. However, prior to the current study, monkeys had not participated in any study of delay of gratification.

In Experiment 2, we also tested four chimpanzees housed at the Language Research Center: Lana (age 40 female), Mercury (age 23 male), Panzee (age 24 female), and Sherman (age 37 male). Chimpanzees were socially housed, but separated into individual 54 m³ enclosures for testing, and like the monkeys they maintained visual and auditory access to other chimpanzees during testing. Chimpanzees had 24-h access to water and were fed various fruits, vegetables, and protein sources following morning and afternoon test sessions as well as at the end of the day (between 1600 and 1800 h). These four individuals had previously participated in multiple delay of gratification tasks of varying degrees of similarity to the tasks used in the present study (Beran 2002; Beran and Evans 2006, 2009; Evans and Beran 2007b).

Apparatus

In Experiments 1 and 2, the capuchin monkeys' test enclosure and apparatus stood 1.07 m above the floor. A vertical Plexiglas panel (33×46 cm) was attached to one end of the test enclosure. A Plexiglas pan (15×7.5 cm) was hinged to the panel so its contents were accessible from both the experimenter and monkey sides of the panel. A dead bolt could be used to lock the pan so only the experimenter had access to its contents.

Chimpanzees were tested in Experiment 2 with a slightly different apparatus (due to the constraints of the different materials that comprised the test enclosures). It consisted of a shallow plastic bowl that was positioned part way through a gap at the floor of the test enclosure so that its contents could be accessed by both the experimenter and the chimpanzee. A second bowl was positioned approximately 1 m from the test enclosure and contained the test items.

Procedure

All testing occurred between 1000 and 1300 h, 2–5 days a week (one session per day). To ensure that animals were similarly motivated to test, we provided each species with a similar sized fraction of their daily diet prior to testing. Capuchin monkeys received one small treat item for voluntarily separating from their group for testing. Chimpanzees were fed one vegetable 30–60 min before testing (chimpanzees required no incentive to separate themselves for testing).

Experiment 1: Capuchin monkey accumulation performance with food and tokens

Testing was conducted in four phases in which we varied the number of items (5 or 50) and type of items (food items: raisins; or tokens: $1 \times 1.5 \times 1.5$ cm plastic toy blocks of various colors) that would accumulate in the Plexiglas pan during trials. In all cases, the experimenter stood approximately 45 cm from the apparatus and hand transferred the items from a 15 cm clear bowl to the test pan, 1 item every 2 s. Test sessions consisted of a mix of forced-accumulation and free-accumulation trials. The only difference between these two trial types was whether the pan was locked, allowing only the experimenter to access the pan's contents, or unlocked, so that both the monkey and the experimenter could access the pan's contents. In forced-accumulation trials, the experimenter transferred every item into the test pan before opening the lock (and thus allowing monkeys access to the items). In free-accumulation trials, the experimenter transferred items into the unlocked food pan until the monkey took the items (and all remaining items then were taken away).

In Phase 1, all trials involved 5 food items. We tested monkeys in eight four-trial sessions, each consisting of one forced-accumulation trial followed by three free-accumulation trials. Because each of our test areas included four side-by-side test enclosures, we presented each trial to each of four monkeys before moving on to the next trial. This created a variable inter-trial interval for a particular monkey, ranging from approximately 45–90 s, depending on the performance of the other three monkeys being tested. This general procedure was also used to test capuchin monkeys in the remaining phases of this experiment.

In Phase 2, all trials involved 50 food items. We tested monkeys in 10 two-trial sessions, each consisting of two free-accumulation trials (and zero forced-accumulation trials). We restricted these sessions to this small number of trials to reduce the possibility that monkeys would become satiated before the end of the test session (given the much larger amount of food that could be obtained in individual trials of this phase).

Phase 3 was similar to Phase 1, except for three key differences. First, five tokens accumulated in the test pan instead of five food items. Second, because tokens were not in themselves inherently rewarding objects, monkeys had to exchange them for individual food items at the end of each trial. Third, because the monkeys were unfamiliar with the tokens prior to participating in this experiment, we conducted a token training phase just prior to beginning Phase 3.

The token training phase consisted of 14-trial sessions in which monkeys chose between one of two token quantities presented in separate 15-cm bowls. We limited the token training phase to just one quantity comparison because we already had independent evidence that these animals could make judgments between arbitrary (i.e., non-edible) stimuli on the basis of their numerosity and correspondence to specific food amounts (e.g., Beran et al. 2008). Token sets were presented on a choice apparatus and monkeys made a choice by reaching through a hole in their test compartment and touching the bowl containing their preferred set of tokens (and the experimenter immediately pulled away the unselected token set). The first two trials of each session were forced-choice trials in which monkeys were presented only a set of three tokens on a randomly selected side of the choice apparatus. The remaining 12 trials were free-choice trials in which monkeys chose between two bowls, one containing three tokens, and another containing one token. In both trials types, monkeys were then given the opportunity to hand out individual tokens through a hole on the side of the test enclosure (i.e., the experimenter held his or her hand, palm up, just below the hole until the monkeys passed out a token). Each token that was handed back to the experimenter earned the monkey one raisin, which was immediately given to the monkey through the same hole on the side of test enclosure. Each monkey participated in these training sessions until they chose the set of three tokens and exchanged them for food items in 10 or more free-choice trials in two consecutive sessions. All of these capuchin monkeys had prior experience exchanging tokens with human experimenters within the context of at least one published or unpublished economic decision making study (Brosnan and de Waal 2003, 2004a, b, 2009; Brosnan et al. 2011; van Wolkenten et al. 2007; also data from Hopper and colleagues, not yet published, and Parrish and Brosnan, not yet published), and so these training sessions only required monkeys to generalize their existing token exchange capacity to the present test environment.

Phase 4 was very similar to Phase 2 with the exception that 50 tokens accumulated in place of 50 food items and monkeys could exchange the tokens for food items at the end of each trial (as in the token training phase).

Experiment 2: Comparing capuchin monkey and chimpanzee accumulation performance with food and tokens

Monkeys were tested using the same general procedure and materials as in Experiment 1. Chimpanzees were tested using a similar procedure, but with the following differences. The experimenter presented items to chimpanzees one at a time every 6 s (the same inter-item interval could not be used with both species, given the greater performance level of the chimpanzees in previous food accumulation tasks, so we instead chose an inter-item interval that was challenging for the chimpanzees in a previous delay maintenance experiment; Beran and Evans 2006). The items presented to chimpanzees were M&M candies and $3 \times 3 \times 6$ cm plastic toy blocks of various colors (the same items could not be used with both species due to differences in daily caloric intake and hand size, respectively, between species). Chimpanzees had previously been trained to associate individual tokens of this type with individual preferred food items in the context of a numerical cognition study (Beran et al. 2011). Chimpanzees exchanged tokens for food items following accumulation trials by carrying tokens (in the bowl) to a section of their enclosure approximately 2 m away from the accumulation site where a second experimenter would receive tokens and pass food items through the wire mesh.

Each test session included two free-accumulation trials each consisting of 20 items. Unlike, in Experiment 1, one trial involved food items and the other involved tokens, and trial order was counterbalanced across sessions. We presented monkeys with 10 sessions. We presented chimpanzees with a minimum of five sessions, and we assessed a performance criterion of three consecutive sessions in which all 20 food items and all 20 tokens were obtained. We presented eight additional sessions to the one chimpanzee that did not meet this criterion.

Data analysis

Because data were not normally distributed, we used non-parametric statistics (Wilcoxon matched-pairs test, Mann–Whitney U test, and Spearman correlation). Given the small sample size, we used the "exact" variant of the above statistical tests (Mundry and Fischer 1998) whenever possible.

Results

Experiment 1: Capuchin monkey accumulation performance with food and tokens

In Phase 1, seven monkeys (all but Drella) accumulated all five items in all three free-accumulation trials of one or

Table 1 Individual capuchin monkey accumulation performance in Experiment 1

Monkey		Mean items accumulated			
Name	Sex	Phase 1: 5 foods	Phase 2: 50 foods	Phase 3: 5 tokens	Phase 4: 50 tokens
Drella	M	1.42	2.38	2.00	1.75
Gabe	M	3.13	5.38	1.71	1.70
Griffin	M	2.75	3.50	3.38	3.05
Liam	M	3.50	9.88	3.00	3.20
Lily	F	3.63	8.13	3.00	6.75
Logan	M	2.67	6.88	4.50	5.50
Nala	F	2.71	4.25	3.79	2.00
Wren	F	4.50	10.94	4.58	4.45
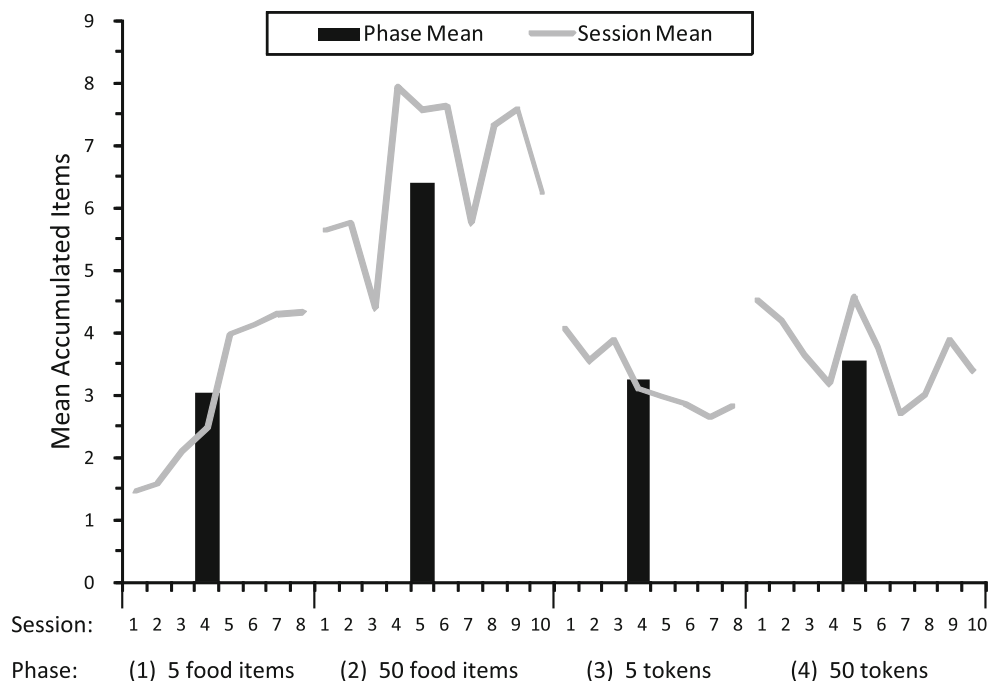
Fig. 1 Mean capuchin monkey accumulation performance in Experiment 1. *Black vertical bars* represent the mean of monkeys' performance in each phase of the experiment. *Gray lines* represent the mean of monkeys' performance over individual sessions of each phase

more sessions (see also Table 1 for individual mean performances). There was a positive correlation between session number and the mean number of items accumulated by the monkeys (Fig. 1; $r_s = 1.0$, $N = 8$, $P < .001$).

In Phase 2, no monkey accumulated all 50 food items in a single trial. However, four monkeys (Liam, Lily, Logan, and Wren) did obtain 10 or more items in multiple trials (see Table 1 for individual mean performances). There was no relationship between session number and the mean number of items accumulated by the monkeys (Fig. 1; $r_s = .378$, $N = 10$, $P = .281$).

In Phase 3, all monkeys met the training criterion for free-choice token trials in 3–7 sessions. Five monkeys (all but Drella, Gabe, and Liam) accumulated all five tokens in all three free-accumulation trials of one or more sessions.

There was no difference between monkeys' performance in this phase and their performance with five food items in Phase 1 (Table 1; $T = 13.5$, $N = 8$, $P = .578$). There was a negative correlation between session number and the mean number of tokens accumulated by the monkeys (Fig. 1; $r_s = -.952$, $N = 8$, $P < .001$).

In Phase 4, no monkey accumulated the full 50 tokens in a single trial. Three monkeys (Logan, Lily, and Wren) each obtained 10 or more items in only a single trial. Monkeys accumulated significantly fewer items in this phase in comparison to Phase 2 in which they were presented with up to 50 food items (Table 1; $T = 0$, $N = 10$, $P = .008$). There was no relationship between session number and the mean number of tokens accumulated by the monkeys (Fig. 1; $r_s = -.467$, $N = 10$, $P = .174$). Given that some

Table 2 Capuchin monkey accumulation performance in Experiment 2

Monkey	Mean items accumulated	
	20 Foods	20 Tokens
Drella	1.20	1.30
Gabe	3.90	1.40
Griffin	3.20	3.00
Liam	6.00	3.90
Lily	14.60	7.90
Logan	9.70	5.40
Nala	2.30	1.50
Wren	9.70	4.90

monkeys had more token experience prior to the present study, we also assessed whether such experience was related to individual differences in token accumulation performance. The three monkeys with the greatest amount of previous experience (Drella, Griffin, and Lily) did not consistently outperform the less experienced monkeys (Table 1) and so there was no relationship between previous token experience and performance in the token accumulation task.

Experiment 2: Capuchin monkey and chimpanzee accumulation performance with food and tokens

Capuchin monkeys accumulated significantly more food items than tokens in this experiment (Table 2; $T = 1.0$, $N = 8$, $P = .016$). One monkey (Lily) accumulated all 20 food items in 5 of 10 trials. No monkey accumulated all 20 tokens in a single trial. There was a positive correlation between session number and the mean number of food items monkeys accumulated (Fig. 2; $r_s = .828$, $N = 10$, $P = .003$). There was no relationship between session number and the mean number of tokens monkeys accumulated (Fig. 2; $r_s = .049$, $N = 10$, $P = .894$).

Given the small number of chimpanzees, we only analyzed their performance within-subjects. Lana, Mercury, and Sherman each obtained all 20 food items and all 20 tokens in 3–5 of their first five sessions, thereby quickly meeting the performance criterion and thus did not require any further sessions (Table 3). Panzee, however, consistently accumulated all 20 food items in her first five sessions, but she obtained less than five tokens during each of those sessions. Consequently, we presented her with 8 additional sessions and her performance remained the same (Table 3; correlation between session number and number of tokens accumulated: $r_s = -.399$, $N = 13$, $P = .177$). Thus, Panzee accumulated significantly more food items than tokens in this experiment ($T = 0$, $N = 13$, $P < .001$), but the other three chimpanzees showed no difference in

performance depending on whether they were watching food items or tokens accumulate.[1]

Discussion

In Experiment 1, capuchin monkeys exhibited some success in delaying gratification for accumulating food items. They gradually learned to accumulate most or all available items in the initial phase involving five available food items. They then took advantage of the greater number of rewards available in the 50 food item phase. In Experiment 2, capuchin monkeys exhibited approximately the same overall level of accumulation performance for food and tokens, providing a clear within-subjects replication despite having added experience with the task, and despite the randomized presentation order of the two trial types. It is interesting to compare this performance to capuchin monkeys tested in previous studies (Addessi et al. submitted; Anderson et al. 2010; Pelé et al. 2011). The monkeys tested here delayed gratification for approximately 3–30 s across the two experiments, and this overlapped with the performance of monkeys tested by Anderson et al. (2010) and Pelé et al. (2011), which delayed gratification for approximately 10–20 and 30–40 s, respectively. Also, in the present study, delay maintenance emerged in the absence of extensive training, contrary to what happened in those two previous studies, whereas the only other study to date that did not include any training showed a more limited evidence of delay maintenance (Mean = 2.71 s; Addessi et al. submitted). Taken altogether, the results of these studies indicate that there is considerable variability in the delay of gratification capacity of this species, and this variability does not appear to be driven by previous training. Rather, it is possible that this apparent continuum of performance is within the normal range of within-subjects variability, given some differences in individual life history. However, future studies will be needed to determine whether there are any particular variables that relate directly to this range of performance.

When we first transitioned monkeys to tokens in place of food items, in Experiment 1, their performance fell gradually across the five token phase and they accumulated significantly fewer items, on average, than in the five food item phase. Then, when given the opportunity to accumulate a greater number of tokens, their performance level

---

[1] Of particular interest were two food accumulation trials with the chimpanzees (one trial each with Sherman and Panzee) in which the experimenter accidentally missed the accumulation bowl and the food item landed on the floor. In both cases, the chimpanzee picked up the food item, but did not eat it, and instead, immediately placed the item into the accumulation bowl. The experimenter continued the trial as usual.
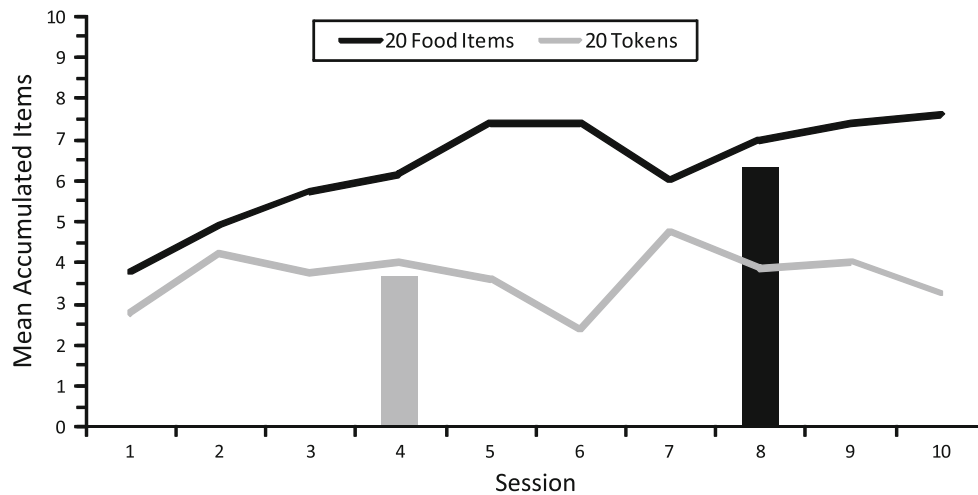
Fig. 2 Mean capuchin monkey accumulation performance in Experiment 2. *Black and gray vertical bars* represent the means of monkeys' performance in trials involving 20 food items and 20 tokens, respec-
tively. *Black and gray lines* represent the means of monkeys' performance over individual sessions involving food and token trials, respectively

remained low. Also, in Experiment 2, monkeys accumulated more food items than tokens and, across consecutive sessions, they did not learn to obtain more tokens. This occurred despite the added experience monkeys gained in the task and despite the change in trial type presentation order from Experiment 1 (i.e., the mixing of trial types within sessions as opposed to separating trial types into different phases). One chimpanzee (Panzee) partially replicated the monkeys' pattern of results, as she accumulated significantly more food items than tokens but did not improve her performance with food items over the course of the experiment (because she was at ceiling). The other three chimpanzees' performance contrasted strongly with the monkeys' performance, as they quickly learned to maximize the accumulation of items regardless of item type and despite experiencing longer inter-item presentation delays. It is important to note that these chimpanzees have years of testing experience in the domain of delayed gratification, so the results of this experiment may not necessarily generalize to other chimpanzees. However, it is equally important to note that these chimpanzees were reported to show excellent delay maintenance skills even in their very first exposure to this task (Beran 2002).

This overall pattern of results consistently shows that tokens do not improve delay maintenance in an accumulation task with either capuchin monkeys or chimpanzees—in fact, in many instances, the use of tokens hampers individual performance. Capuchin monkeys delayed gratification significantly more for food items than for tokens in both experiments, and one chimpanzee also showed this pattern, whereas the others showed no difference in performance between accumulating tokens and food items. The inefficacy of tokens in the accumulation task suggests that the

prepotent response requiring inhibition is triggered by quantity rather than by the immediate presence of food rewards. This is consistent with the lack of effect of low-symbolic tokens in facilitating performance in the reverse-reward contingency task (Addessi and Rossi 2011; Boysen 2006), and constitutes an important finding on the mechanisms responsible for failures of delay maintenance. Unless the quantitative dimension of the tempting stimulus is abstracted away, symbolic tokens do not facilitate self-control. Such increased abstraction might be hard to implement in the accumulation task, because the accumulating rewards need to be discrete: This is probably what makes this task so challenging, even for animals that demonstrate remarkable tolerance for delay in other self-control paradigms (for a discussion, see Addessi et al. submitted).

However, this is the first time that symbolic stimuli have been reported to hurt performance when used to represent food rewards in a self-control test, as opposed to just being ineffective. This may have occurred because of the difference in time that was required to obtain and consume food items in the token and food versions of the accumulation task. We originally designed these two variations of the task so that, given a particular amount of test items, nearly the same amount of time would elapse between the beginning of the trial and the moment in which a monkey could *begin* consuming the rewards, regardless of the test item type. After all, this is the critical moment in any delay maintenance task. This duration was virtually the same for the token and food item versions of the task because the amount of time that a monkey required to exchange a single token for a food item was negligible (in fact monkeys often immediately pushed the first token out of the side of the test enclosure, presumably in anticipation of receiving their first

Table 3 Chimpanzee accumulation performance in Experiment 2

Chimpanzee	Session	Mean items accumulated	
		20 Food items	20 Tokens
Lana	1	20	20
	2	20	20
	3	20	20
	4	20	20
	5	20	20
Mercury	1	20	3
	2	20	2
	3	20	20
	4	20	20
	5	20	20
Panzee	1	20	4
	2	20	2
	3	20	2
	4	20	2
	5	20	4
	6	20	1
	7	20	2
	8	20	3
	9	20	2
	10	20	2
	11	20	2
	12	20	1
	13	20	2
Sherman	1	20	20
	2	20	20
	3	20	20
	4	20	20
	5	18	20

food reward after ending the trial). What this design element did not account for, however, was the *total* duration that would be required to obtain a given number of food rewards via token exchange. Because approximately 1–2 s were required to exchange each token for a reward, the amount of delay associated with a token accumulation trial of a given number of items could potentially be twice the delay experienced in an equivalent food item accumulation trial. Therefore, it may not be a coincidence that capuchins accumulated approximately twice as many test items in food item trials than in token trials. This suggests the need for additional studies that would allow for all of the corresponding food items to be delivered at once as soon as the animal touches the accumulated items.

In sum, this study provides the first demonstration that tokens having a one-to-one correspondence with food items are ineffective to facilitate delay maintenance in an accumulation task, both with capuchin monkeys and chimpanzees, thus corroborating the hypothesis that prepotent responses in such situations are driven by quantity rather than by the mere presence of a primary reinforcer. In fact, symbolic tokens even decrease performance for some individuals, possibly due to the additional delay introduced by token exchange. Thus, it remains important to further probe instances where tokens and other forms of abstract representation may facilitate types of behavioral inhibition toward the end of maximizing reward and obtaining the best possible outcomes. Finally, in this study, most chimpanzees continued to display excellent self-control capacity with respect to the accumulation paradigm, as they have in all previous reports, with a performance qualitatively beyond that of capuchin monkeys. This suggests a need for testing with additional species in these paradigms to further illuminate the spectrum of delay of gratification capacity in non-human animals.

References

Addessi E, Rossi S (2011) Tokens improve capuchin performance in the reverse–reward contingency task. Proc R Soc B Biol Sci 278:849–854. doi:10.1098/rspb.2010.1602

Addessi E, Paglieri F, Focaroli V (2011) The ecological rationality of delay tolerance: insights from capuchin monkeys. Cognition 119:142–147

Anderson JR, Kuroshima H, Fujita K (2010) Delay of gratification in capuchin monkeys (*Cebus apella*) and squirrel monkeys (*Saimiri sciureus*). J Comp Psychol 124:205–210. doi:10.1037/a0018240

Baumeister RF, Vohs KD (eds) (2004) Handbook of self-regulation: research theory and applications. Guilford Press, New York

Beran MJ (2002) Maintenance of self-imposed delay of gratification by four chimpanzees (*Pan troglodytes*) and an orangutan (*Pongo pygmaeus*). J Gen Psychol 129:49–66. doi:10.1080/0022130020 9602032

Beran MJ, Evans TA (2006) Maintenance of delay of gratification by four chimpanzees (*Pan troglodytes*): the effects of delayed reward visibility, experimenter presence, and extended delay intervals. Behav Process 73:315–324. doi:10.1016/j.beproc.2006.07.005

Beran MJ, Evans TA (2009) Delay of gratification by chimpanzees (*Pan troglodytes*) in working and waiting situations. Behav Process 80:177–181. doi:10.1016/j.beproc.2008.11.008

Beran MJ, Savage-Rumbaugh ES, Pate JL, Rumbaugh DM (1999) Delay of gratification in chimpanzees (*Pan troglodytes*). Dev Psychobiol 34:119–127. doi:10.1002/(SICI)1098-2302(199903)34: 2<119:AID-DEV5>3.0.CO;2-P

Beran MJ, Harris EH, Evans TA, Klein ED, Chan B, Flemming TM, Washburn DA (2008) Ordinal judgments of symbolic stimuli by capuchin monkeys (*Cebus apella*) and rhesus monkeys (*Macaca mulatta*): the effects of differential and nondifferential reward. J Comp Psychol 122:52–61

Beran MJ, Evans TA, Hoyle D (2011) Numerical judgments by chimpanzees (*Pan troglodytes*) in a token economy. J Exp Psychol Anim Behav Process 37:165–174. doi:10.1037/a0021472

Boysen ST (2006) The impact of symbolic representations on chimpanzee cognition. In: Hurley S, Nudds M (eds) Rational animals?. Oxford University Press, New York, pp 489–511

Boysen ST, Bernston GG (1995) Responses to quantity: perceptual versus cognitive mechanisms in chimpanzees. J Exp Psychol Anim Behav Process 21:82–86

Boysen ST, Mukobi KL, Bernston GG (1999) Overcoming response bias using symbolic representations of number by chimpanzees (*Pan troglodytes*). Animal Learn Behav 27:229–235

Brosnan SF, de Waal FBM (2003) Monkeys reject unequal pay. Nature 425:297–299

Brosnan SF, de Waal FBM (2004a) A concept of value during experimental exchange in brown capuchin monkeys, Cebus apella. Folia Primatol 75:317–330. doi:10.1159/000080209

Brosnan SF, de Waal FBM (2004b) Socially learned preferences for differentially rewarded tokens in the brown capuchin monkey (*Cebus apella*). J Comp Psychol 118:133–139

Brosnan S, de Waal F (2009) *Cebus apella* tolerate intermittent unreliability in human experimenters. Int J Primatol 30(5):663–674. doi:10.1007/s10764-009-9366-x

Brosnan SF, Parrish A, Beran MJ, Flemming T, Heimbauer L, Talbot CF, Lambeth SP, Schapiro SJ, Wilson BJ (2011) Responses to the assurance game in monkeys, apes, and humans using equivalent procedures. Proc Nat Acad Sci 108:3442–3447. doi:10.1073/pnas.1016269108

Chelonis JJ, King G, Logue AW, Tobin H (1994) The effect of variable delays on self-control. J Exp Anal Behav 62:33–43. doi:10.1901/jeab.1994.62-33

Cheng K, Peña J, Porter M, Irwin J (2002) Self-control in honeybees. Psychon Bull Rev 9:259–263. doi:10.3758/bf03196280

Evans TA, Beran MJ (2007a) Chimpanzees use self-distraction to cope with impulsivity. Biol Lett 3:599–602. doi:10.1098/rsbl.2007.0399

Evans TA, Beran MJ (2007b) Delay of gratification and delay maintenance in rhesus macaques (*Macaca mulatta*). J Gen Psychol 134:199–216

Green L, Estle SJ (2003) Preference reversals with food and water reinforcers in rats. J Exp Anal Behav 79:233–242

Green L, Fry AF, Myerson J (1994) Discounting of delayed rewards: a life-span comparison. Psychol Sci 5:33–36. doi:10.1111/j.1467-9280.1994.tb00610.x

Grosch J, Neuringer A (1981) Self-control in pigeons under the Mischel paradigm. J Exp Anal Behav 35:3–21. doi:10.1901/jeab.1981.35-3

Hopper LM, Essler J, Freeman H, Talbot CF, Lambeth SP, Schapiro SJ, Williams LE, Brosnan SF (in prep) Enough is enough: the effect of the frequency of unequal outcomes on decisions to accept food rewards in four primate species

Killeen PR, Smith JP, Hanson SJ (1981) Central place foraging in Rattus norvegicus. Anim Behav 29:64–70. doi:10.1016/s0003-3472(81)80152-2

Lawyer SR, Williams SA, Prihodova T, Rollins JD, Lester AC (2010) Probability and delay discounting of hypothetical sexual outcomes. Behav Process 84:687–692. doi:10.1016/j.beproc.2010.04.002

Logue AW (1988) Research on self-control: an integrating framework. Behav Brain Sci 11:665–709. doi:10.1017/S0140525X00053978

Logue AW, Pena-Correal TE (1985) The effect of food deprivation on self-control. Behav Process 10:355–368

Logue AW, Forzano LB, Ackerman KT (1996) Self-control in children: age, preference for reinforcer amount and delay, and language ability. Learn Motiv 27:260–277. doi:10.1006/lmot.1996.0014

Mazur JE (2007) Species differences between rats and pigeons in choices with probabilistic and delayed reinforcers. Behav Process 75:220–224

Mischel W (1974) Processes in delay of gratification. Adv Exp Soc Psychol 7:249–292. doi:10.1016/S0065-2601(08)60039-8

Mischel W, Shoda Y, Rodriguez ML (1989) Delay of gratification in children. Science 244:933–938

Mundry R, Fischer J (1998) Use of statistical programs for nonparametric tests of small samples often leads to incorrect P values: examples from animal behaviour. Anim Behav 56:256–259

Parrish A, Brosnan SF (in prep) Changing the relative difference in outcomes affects reactions to unequal rewards in capuchin monkeys (*Cebus apella*)

Pelé M, Micheletta J, Uhlrich P, Thierry B, Dufour V (2011) Delay maintenance in tonkean macaques (*Macaca tonkean*) and brown capuchin monkeys (*Cebus apella*). Int J Primatol 32:149–166. doi:10.1007/s10764-010-9446-y

Stevens JR, Hallinan EV, Hauser MD (2005) The ecology and evolution of patience in two new world monkeys. Biol Lett 1:223–226. doi:10.1098/rsbl.2004.0285

Stevens JR, Rosati AG, Heilbronner SR, Mühlhoff N (2011) Waiting for grapes: expectancy and delayed gratification in bonobos. Int J Comp Psychol 24:99–111

Tobin H, Chelonis JJ, Logue AW (1993) Choice in self-control paradigms using rats. Psychol Rec 43:441–453

Tobin H, Logue AW, Chelonis JJ, Ackerman KT, May JGI (1996) Self-control in the monkey *Macaca Fascicularis*. Animal Learn Behav 24:168–174

Toner IJ, Smith RA (1977) Age and overt verbalization in delay-maintenance behavior in children. J Exp Child Psychol 24:123–128. doi:10.1016/0022-0965(77)90025-X

van Haaren F, van Hest A, van de Poll NE (1988) Self-control in male and female rats. J Exp Anal Behav 49:201–211. doi:10.1901/jeab.1988.49-201

van Wolkenten M, Brosnan SF, de Waal FBM (2007) Inequity responses of monkeys modified by effort. Proc Nat Acad Sci 104:18854–18859. doi:10.1073/pnas.0707182104

Vick S-J, Bovet D, Anderson J (2010) How do African grey parrots (*Psittacus erithacus*) perform on a delay of gratification task? Animal Cognit 13:351–358. doi:10.1007/s10071-009-0284-2

Vlamings PHJM, Uher J, Call J (2006) How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: the effects of food quantity and food visibility. J Exp Psychol Anim Behav Process 32:60–70. doi:10.1037/0097-7403.32.1.60