



Familiarity, Active Memory And Pressure During A Delayed Matching-To-Sample Task In Tufted Capuchin Monkeys (*Sapajus [Cebus] Apella*)

Meghan J. Sosnowski^{1,2,3,*} and Sarah F. Brosnan^{1,2,4,5}

¹ Department of Psychology, Georgia State University

² Language Research Center, Georgia State University

³ California National Primate Research Center, University of California, Davis

⁴ Center for Behavioral Neuroscience, Georgia State University

⁵ Neuroscience Institute, Georgia State University

*Corresponding author (Email: meg.sosnowski@gmail.com)

Citation – Sosnowski, M. J., & Brosnan, S. F. (2025). Familiarity, active memory and pressure during a delayed matching-to-sample task in tufted capuchin monkeys (*Sapajus [Cebus] Apella*). *Animal Behavior and Cognition*, 12(1), 32-44. <https://doi.org/10.26451/abc.12.01.02.2025>

Abstract – Delayed matching-to-sample (DMTS) tasks are commonly used in the field of comparative cognition to study memory, including working memory. However, specific task demands vary across studies and species, and as such, DMTS tasks may engage different memory systems when features such as the available stimulus pool differ. Further, individual or species-wide differences in response to pressure to perform may increase variation within a species. We explored how task features, memory systems, and pressure interact in tufted capuchin monkeys (*Sapajus [Cebus] apella*) to influence performance on a DTMS task by varying the size of the possible stimulus pool across testing blocks. We also varied the amount of pressure within a testing block by training monkeys to associate a background color change with a more difficult, but more highly rewarded, trial, as we had done in previous work. In accordance with previous literature (Basile & Hampton, 2013), we found that performance greatly decreased when the possible stimulus pool was limited as compared to a large possible stimulus pool, likely because monkeys could not rely on passive familiarity memory to complete the task. However, we found no overall species tendency to fail under pressure in either the limited- or large-set conditions; instead, we found a surprising tendency to thrive under higher pressure. Taken together, our results further highlight the importance of considering DMTS task features when studying specific memory systems in non-human species and suggest that the DTMS task might not be the best paradigm for testing pressure effects without consideration of individual differences.

Keywords – Primate, Cognition, Delayed Matching-To-Sample, DMTS, Pressure

Humans and non-human animals regularly face situations in which they must perform at their best or make a good decision to ensure a positive outcome and avoid a negative one; these situations, in which an individual's outcome is positive or negative based on their own performance, are said to have an element of pressure. Until recently, pressure has only been studied in humans. However, we have good reason to expect that other species would also be susceptible to effects of high-pressure situations. First, other species show similar cognitive abilities to those that, in humans, are particularly prone to interference from pressure (for instance, working memory-like systems, Roberts, 1996). Further, high-pressure situations, in which outcomes as important as survival are tied to decisions, occur frequently in the wild; therefore, it seems likely that the ability to respond to pressure appropriately would have conferred an evolutionary benefit.

Despite this, only a few studies have specifically explored how pressure influences decision-making in non-humans. Understanding this will broaden our current knowledge about the effects of pressure beyond human subjects and provide a comparative perspective to when such effects occur and how individuals have evolved to ameliorate them.

Most of this previous work in pressure has focused on two cognitive systems that seem to be affected by pressure – procedural memory and working memory (a cognitive system which information is temporarily held, manipulated, and updated prior to being stored in long-term memory). Procedural memory, which is the type of memory involved in pitching a baseball, riding a bike, or completing a task using a well-known sequence, has been a staple of explorations of choking under pressure in humans, probably because it is the memory system that is most engaged when performing in the domain most associated with choking under pressure: sports. Expert golfers sometimes choke when attempting an important putt (Beilock et al., 2002), soccer strikers are estimated to choke on 7% of penalty kicks (Dohmen, 2008), and collegiate level baseball batters drastically change their usual swing (often to negative effects) when the pressure is on (Gray, 2004). However, negative effects of pressure are not limited to physical actions, and choking occurs in other types of procedural and working memory processes, such as mathematical calculation (Beilock, 2008; Beilock et al., 2004), especially when participants tend to use more complex working-memory strategies (Beilock & Carr, 2005; Beilock & DeCaro, 2007; Mattarella-Micke et al., 2011). It therefore seems as though the additional working memory load associated with pressure when performing may interfere with the ability to complete tasks and procedures, at least in humans.

Although other animals show evidence of working memory-like systems and face pressure-ridden situations in their natural lives, there is a scarcity of research about how directly manipulating pressure in the absence of a difficulty difference affects performance and decision-making in non-human species (for a review, see Sosnowski & Brosnan, 2023). Those studies that have manipulated task demand with the intent to study pressure specifically were exclusively focused on non-human primates, presumably because these other primate species show evidence of similar cognitive systems to those susceptible to pressure in humans (for instance, working memory, see Roberts, 1996; or task sequencing and monitoring, see Desrochers et al., 2016). Two studies have attempted to specifically isolate the effect of pressure on performance, and each focused on a different cognitive system that shows such effects in humans. In the first, Smoulder et al. (2021) tested three rhesus macaque subjects (*Macaca mulatta*) on a psychomotor learning task in which they needed to reach their arm toward an on-screen target, while reaction times, fine motor adjustments, and success or failure to reach toward the correct area were recorded. They found that the monkeys were likely to fail on trials which were far more highly rewarded in comparison to typical trials, suggesting that the monkeys were susceptible to pressure induced by a particularly valuable reward.

Concurrently, we explored pressure in a different cognitive system (memory) in a different species (tufted capuchin monkeys: *Sapajus [Cebus] apella*; Sosnowski et al., 2022). We trained the monkeys to complete a delayed-matching-to-sample (DMTS) task in which they first saw a sample stimulus that disappeared after two seconds and then, after a brief delay, the monkeys had to select the sample out of an array of four choice stimuli. During training, we also trained them to associate a change in background color with a more difficult, but more highly-rewarded, trial that featured a significantly longer delay than a typical trial (five seconds as compared to one second); these higher-pressure trials also featured a greater consequence for an incorrect answer (a longer inter-trial interval period, or ITI, during which the monkeys could not work on the computer). During testing, however, we removed the difference in difficulty between higher-pressure and typical trials (both the delay and the ITI) while maintaining the differences in appearance and the outcomes of the higher-pressure trials. Thus, any differences between the higher-pressure and typical trials in testing should be solely due to the effect of pressure to perform, without the confound of differences in difficulty.

We found that there was significant individual variation in choking or thriving under pressure, as we see in humans, with some monkeys performing significantly worse on higher-pressure trials and some, counterintuitively, performing better, indicating that the monkeys were sensitive to the differential pressure conditions (Sosnowski et al., 2022). Although both this study and the one conducted by Smoulder and

colleagues (2021) suggest that at least some other primate species (from both the *Platyrrhini* and *Catarrhini* parvorders) are sensitive to and sometimes respond negatively to pressure, important questions remain about the systems affected by pressure, the contexts under which individuals choke or thrive, and the individual differences associated with performance. Moreover, in our previous study, we found no overall effect, and predicted that we might find a stronger species-level effect when subjects were reliant on a memory system that has, in humans, been linked to stronger pressure effects: working memory. Thus, in the present study, we use the same task that we did in our previous work, the DMTS, but with the goal of differentiating between two systems that might be engaged by the task based on specific task demands, one based on familiarity and one on working memory.

DMTS tasks have been previously used as a measure of working memory in this species (Tavares & Tomaz, 2002), but because the monkeys may not need to engage or modify the information while they are holding the sample in memory, an argument can be made that most DMTS tasks can be successfully completed using solely short-term memory or familiarity memory. Familiarity memory, or the ability to recognize stimuli, individuals or events that have been previously encountered, notably does not require active recollection of the event in question, making it a much less active form of memory than working memory (and thus, potentially less susceptible to interference from pressure). Importantly, many cognitive tasks can be completed using multiple strategies, which might even use different cognitive systems, and we cannot assume that different species use the same strategies or systems when completing these tasks. For instance, given a computerized task in which human and rhesus macaque subjects were asked if a given image was the same or different to a previously presented target from earlier in a sequence, monkeys showed evidence of using a non-selective memory strategy (that is, one reliant on recency or familiarity) rather than target-selective memory strategies used by humans (Wittig et al., 2016). Thus, it's important to consider that the cognitive system that a task is designed to engage may not actually be the one being used, and instead species may find simpler ways to complete the tasks.

Indeed, even within a species, specific task features might drive which cognitive system is engaged during testing. There is evidence from rhesus macaques that the number and familiarity of available stimuli affect which type of memory is engaged when completing a DMTS task (Basile & Hampton, 2013) – namely, if there is a large pool of unfamiliar potential stimuli used in the DMTS, monkeys can use familiarity memory to solve the task, as any stimulus that is particularly familiar is almost certainly the sample that was shown just a few seconds before. However, if the same four stimuli are presented for every trial, then all four stimuli become familiar. In this case, an individual must actively remember which specific stimulus was just presented as a sample within the present trial – requiring some element of active memory maintenance. In Basile and Hampton's study, this distinction was demonstrated by a difference in performance as the result of an intermediary task only in the latter type of trial, while the former type of trial, in which familiarity was a viable strategy, remained unaffected (2013). This methodology of manipulating the number of available stimuli was also tested in orangutans, but orangutans seemed to engage working memory even when it wasn't necessary, demonstrating that the task could be used to assess cognitive strategies (Brady et al., 2021).

In consideration of these ideas in interpreting our previous study (Sosnowski et al., 2022), the test sessions used a stimulus pool of eight stimuli, of which four were presented on any given trial; therefore, although we maintained consistency in type of stimuli and color (basic black shapes) in order to avoid having the monkeys use simple familiarity to complete the task, the fact that there were more stimuli in the available pool than were presented on any given trial means that monkeys could still rely at least in part of familiarity. Thus, it might be that such a task did not adequately engage working memory enough in some animals to be susceptible to interference by pressure.

The present study uses a similar methodology to differentiate between familiarity memory and working memory in tufted capuchin monkeys, while also adding a manipulation designed to vary the amount of pressure on a trial level. As in our previous work, we trained the monkeys to associate a background color change with a harder, but more highly rewarded, DMTS trial to train them to associate the cue with higher pressure. Then, during testing, we removed this difference in difficulty while having the monkeys complete two different types of DMTS sessions – one large-set type in which there was a very

large pool of potential stimuli from which the sample and distractors were randomly selected, and one limited-set type in which the same four stimuli were used for every trial, with the sample image randomly selected each trial. If, as Basile and Hampton (2013) suggest, monkeys can solve the former type of trials using familiarity memory, while the limited-set trials require working memory engagement, we should see a difference in performance between limited-set and large-set trials. Further, if working memory is more susceptible to pressure effects than familiarity memory in capuchins, we should find that monkeys were more likely to fail on higher-pressure trials in the limited-set trials, which should engage working memory, than in the large-set trials. If, however, both familiarity memory and working memory are susceptible to pressure effects (or conversely, that neither are), then there should be no difference between the sets.

Methods

Ethics Statement

All data collection was approved by the Georgia State University IACUC (#A20018) and was in accordance with the applicable and relevant legal requirements for animal research conducted in the United States. Additionally, all testing procedures adhered to the American Society of Primatologists Principles for the Ethical Treatment of Non-human Primates and the Animal Behavior Society Guidelines for the Ethical Treatment of Non-human Animals in Research.

Subjects and Study Site

We tested 18 tufted capuchin monkeys (*Sapajus [Cebus] apella*; 14 females, 4 males; age range: 11 – 26 years old) residing at the Language Research Center (LRC) of Georgia State University, USA. At the LRC, monkeys lived in one of five mixed-sex social groups ranging in size from four to eight individuals, with the exception of one bachelor pair of males. All groups, including the bachelor group, lived in their own large indoor-outdoor enclosure from which they had visual and auditory access to other groups of capuchins. Each group had access to both the indoor and outdoor components of their enclosure except when they voluntarily chose to separate from their groupmates to participate in testing, for maintenance, or due to inclement weather. Subjects were previously trained using positive reinforcement methods to voluntarily enter individual testing chambers for the purposes of testing. Monkeys could choose not to participate at any time, and there were no consequences for not separating except for not being able to participate in the day's cognitive or behavioral testing. Monkeys were never deprived of food, water, social access or outdoor access for the purpose of encouraging participation in testing. All monkeys had access to food several times per day regardless of their participation. Monkeys had *ad libitum* access to water at all times, including during cognitive testing.

Computerized Testing Apparatus

All capuchin monkeys were previously familiar with the LRC Computerized Testing System (Evans et al., 2008), which consisted of a desktop or laptop computer running a Windows OS connected to a monitor at the appropriate height to be seen from the clear faceplate of the monkeys' individual testing chamber. The computer was also connected to an automatic pellet dispenser, which allowed 45 mg banana-flavored pellets (Bio-Serv) to be distributed as rewards by the experimental programs. Monkeys had been trained to interact with the computer using a modified USB joystick to control a cursor on the screen, allowing them to participate in cognitive tasks. The delayed-match-to-sample task used in the present study was coded and run using Python 3.9 (Python Software Foundation, 2016) and the *pygame* module (Shinners, 2011). We used ClipArt during the training phase, as in our previous work (total training stimuli, $n = 500$), and a different set of small pixel-art images as stimuli throughout testing (total number of images available during testing, $n = 714$). The program recorded subject name, date, time of trial, trial number, all

stimuli used within that trial, type of session, pressure condition of each trial, and success or failure for each trial.

Delayed Matching-to-Sample Task – Training

In order to assess the effect of pressure on different types of memory, we used a DMTS task that varied the size of the available stimulus pool and the pressure conditions. Monkeys began a trial by moving their cursor to a start button. Once the cursor contacted the start button, monkeys were shown a sample image for two seconds, after which time the sample disappeared. Monkeys then underwent a brief delay period (that varied as a function of the pressure condition), after which an array of four stimuli, including the sample, appeared in randomized locations on the screen. Monkeys were then tasked with selecting the stimulus that they saw earlier in that trial out of this array. If they did so correctly, they heard an auditory chime and received a food reward consisting of one or more pellets (depending on trial type). If they did not do so correctly, they heard a buzzer and underwent a longer inter-trial interval as a consequence (described below).

During the training phase of the study, we trained monkeys to associate a background color change with a harder, more highly rewarded trial – a situation that was designed to introduce an element of pressure. On these higher-pressure trials, once monkeys’ cursors collided with the start rectangle, the background of the program immediately changed to blue. Monkeys were then presented with the sample for two seconds, but then underwent a long delay of five seconds prior to making their choice; in our previous work (Sosnowski et al., 2022), this delay was difficult enough to induce differences in performance while still keeping the task approachable and the monkeys engaged. If the monkey was able to complete this higher-pressure trial successfully, an auditory chime played, they received a higher-than-typical payout of five pellets, and the next trial immediately began. If they did not do so successfully, they heard a buzzer and received no pellets; in addition, the program added a longer inter-trial interval (five seconds as compared to two seconds) intended to serve as a “timeout” prior to the next trial as a consequence for the incorrect response.

Higher-pressure trials in training occurred on one out of every five trials, in order to keep the higher-pressure trials at least somewhat rare, as informed by Smoulder and colleagues (2021). The remaining four out of five trials consisted of lower-pressure trials, in which the background remained white after the monkey started the trial by moving their cursor to the start button. Monkeys underwent a delay of only one second prior to making their choice in these lower-pressure trials, the reward for successful choices was a more typical reward of a single banana-flavored pellet, and monkeys underwent a two second inter-trial interval even when they got the trial wrong. Monkeys trained on the task until they were able to achieve 75% correct on each type of trial, after which they were moved onto the testing phase.

Delayed Matching-to-Sample Task – Testing

Trials in the testing phase appeared very similar to those in training, with one key difference. In the testing phase, we removed the differential difficulty on higher-pressure trials while maintaining the visual cues and outcomes that were reinforced in training. In other words, while higher-pressure test trials visually looked very similar to those in training, in testing the delay in these trials was only one second, matching the delay in the lower-pressure trials. Therefore, in testing, any differences between the higher- and lower-pressure trials should be solely the result of the pressure condition, without the confounding factor of differential difficulty. We also used an entirely new set of stimuli comprised of pixel-art images.

Testing Conditions

In order to explore how differing memory processes and strategies might be influenced by pressure, there were two types of sessions in the testing phase that varied in the number of available images to be used as stimuli in the trials, as informed by previous literature (Basile & Hampton, 2013). In large-set, or

familiarity sessions, the program would select four images at random from a large pool of pixel-art images for a given trial; one of these images would be used as the sample while the other three would be used as distractor images in the choice array. Because there were a large number of images (total $n = 714$) available to be chosen for each trial, the likelihood of a given stimulus appearing on two trials in a row was very low (about 0.003%). Therefore, in these sessions, monkeys were able to use familiarity memory to successfully complete the task – images in the choice array that were saliently familiar were likely to be the sample that they just saw at the beginning of the trial, and therefore the correct answer for that trial.

The other type of session, the *limited-set or working memory sessions*, limited the number of stimuli available to be used within each trial to four, such that every trial featured the same four stimuli in the choice array. From these four stimuli, the sample was still selected randomly within each trial, and all stimuli locations were randomized across each trial such that no stimulus was associated with a particular location. However, because the same four stimuli were presented on each trial, familiarity memory was no longer a viable strategy for correctly answering trials beyond the first or second trial; monkeys instead needed to monitor which stimulus had been presented within that specific trial. Previous literature indicated that these types of trials engage working memory rather than familiarity memory (Basile & Hampton, 2013).

Monkeys completed four 250 trial blocks that alternated in session type (limited-set or large-set) across blocks, such that monkeys completed all 250 trials across a block before being switched to the next block and next session type. Monkeys were allowed to complete as many trials as they chose to within a testing day, so the number of testing days varied across monkeys ($Mean_{TestingDays} = 5.17$, $SD_{TestingDays} = 3.22$). To account for order effects, we counterbalanced the first session type experienced across monkeys and accounted for this difference in our statistical analysis.

Statistical Analysis

We first wanted to assess if there was a difference in performance on the typical, or lower-pressure, trials based on the session type to see if these trials indeed seemed to tap into different memory processes, as supported by data from rhesus macaques (Basile & Hampton, 2013) and orangutans (Brady et al., 2021). We did so at a trial level by fitting a simple logistic mixed model predicting each trial's outcome (success or failure) from the fixed effect of block type. We also assessed how experience with the task might differ by including the covariate of block number as a fixed effect. We included subject identity nested within the first block type experienced as a random effect term, to account for our repeated measures design. Because we were interested in assessing the session type manipulation's effect on typical performance, essentially as a manipulation check, in this first model we only included trials that were lower pressure.

Then, to assess if pressure impacted performance differently in the working memory trials as compared to the familiarity trials, we built another logistic mixed-model using the full dataset that included both higher- and lower-pressure trials. We again predicted each trial's outcome (success or failure) using block type, pressure condition, and the interaction between the two as fixed effect predictors. In this model, we again included subject nested within first block type as a random effect term to account for repeated measures. Finally, to explore the effect of pressure in each session type (limited- or large-set) alone, we subset the dataset into limited-set only trials and large-set only trials. For each of these two datasets, we built a separate logistic mixed model predicting trial outcome from pressure condition, while including subject and first block type experienced as a random effect term as before.

We compared each of these models to a corresponding null model consisting only of the intercept and random effect term using a likelihood test. We also assessed the predictive value of each fixed effect variable using Wald C^2 Type II tests. We built models using the *base*, *lme4*, *lmtest*, and *glmer* packages of R 3.6.0 in RStudio (Bates et al., 2015; R Development Core Team, 2016). Predictors were compared using the “Anova” function of the *car* package and likelihood tests were conducted using the “anova” function of the *stats* package.

Results

At the end of testing, we had data from 1000 trials from each of 18 subjects, resulting in an overall dataset of 18000 trials; all monkeys that began testing completed all four blocks (two in the large-set condition and two in the limited-set condition, alternating for each monkey). Descriptive statistics of each individual’s performance in each block type and pressure condition is provided in Table S1.

Monkeys were significantly less likely to succeed on limited-set trials than on large-set trials (*Odds Ratio* = 0.45, *SE* = 0.02, *95% CI* = [0.42 – 0.48], *z* = -22.86, *p* < .001; Table 1) on the typical, lower-pressure trials alone.

Table 1

Logistic Mixed-Model Comparing Performance On Limited-Set Trials To Performance On Large-Set Trials On Lower-Pressure Trials Only, With Block Number Included As A Fixed Covariate

Predictors	Likelihood of Success			
	Odds Ratio	SE	95% CI	<i>p</i>
(Intercept)	1.92	0.18	1.61 – 2.31	< . 001
Block Type (Limited-Set)	0.45	0.02	0.42 – 0.48	< . 001
Block	1.01	0.02	0.98 – 1.04	.560
N _{Subject}	18			
Marginal R ² / Conditional R ²	0.046/ NA			

Note. Full vs. null-model $\chi^2(2) = 535.06, p < .001$. For the categorical predictor of block type, the referent category is large set. Significant estimates and their *p*-values are bolded.

Although there was a slight trend of improvement across blocks, there was not a statistically significant effect of block on likelihood of success on a trial across all blocks (*Odds Ratio* = 1.01, *SE* = 0.02, *95% CI* = [0.98 – 1.04], *z* = 0.58, *p* = .560; Table 1). This pattern held in the overall dataset, which included both higher- and lower-pressure trials; monkeys were much less likely to succeed in limited-set trials than large-set trials (*Odds Ratio* = 0.45, *SE* = 0.02, *95% CI* = [0.42 – 0.48], *z* = -22.79, *p* < .001; Table 2).

Table 2

Logistic Mixed Model Predicting Likelihood Of Success Based On Trial Pressure Condition And Block Type

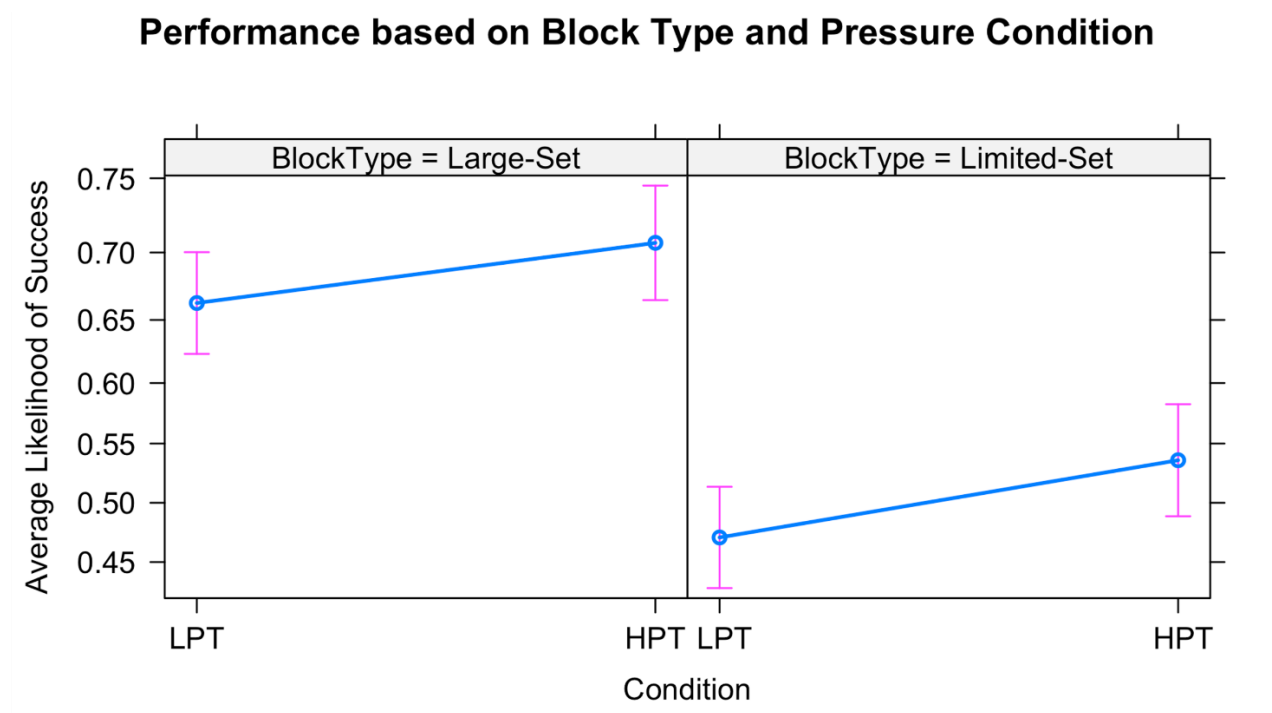
Predictors	Likelihood of Success			
	Odds Ratio	SE	95% CI	<i>p</i>
(Intercept)	1.97	0.17	1.66 – 2.34	< . 001
Block Type (Limited-Set)	0.45	0.02	0.42 – 0.48	< . 001
Pressure Condition (HPT)	1.23	0.07	1.09 – 1.37	< . 001
Block Type * Pressure Condition	1.06	0.08	0.91 – 1.24	.463
N _{Subject}	18			
Marginal R ² / Conditional R ²	0.047/ 0.047			

Note. Full vs. null-model $\chi^2(3) = 675.85, p < .001$. For the categorical predictor of pressure condition the referent category is lower-pressure; for the categorical predictor of block type, the referent is large-set. Significant estimates and their *p*-values are bolded.

However, monkeys were not significantly more likely to fail on higher-pressure trials than on lower-pressure, typical trials; instead, although monkeys counterintuitively performed slightly better on higher-pressure trials overall (*Odds Ratio* = 1.23, *SE* = 0.07, *95% CI* = [1.09 – 1.37], *z* = 3.53, *p* < .001), there was no interaction between session type and pressure condition (*Odds Ratio* = 1.06, *SE* = 0.08, *95% CI* = [0.91 – 1.24], *z* = 0.73, *p* = .463), indicating that monkeys did not perform differently on higher-pressure trials than lower-pressure ones between the two session types (Figure 1).

Figure 1

Effect Plot Of Trial Performance Based On Block Type And Pressure Condition



Note. Lower-Pressure Trial = “LPT”; Higher-Pressure Trial = “HPT”. For a four-choice task, chance performance is a 0.25 likelihood of success; as the y-axis does not go to zero, chance is below the range represented in the graph. Error bars represent standard error of the mean (SE).

This effect held when analyzing the two session types separately. Monkeys were significantly more likely to succeed on higher-pressure trials as compared to lower pressure ones in the limited-set trials (*Odds Ratio* = 1.30, *SE* = 0.07, *95% CI* = [1.17 – 1.44], *z* = 4.89, *p* < .001; Table 3a) as well as in large-set trials (*Odds Ratio* = 1.23, *SE* = 0.07, *95% CI* = [1.10 – 1.37], *z* = 3.53, *p* < .001; Table 3b).

Table 3

Logistic Mixed-Models Predicting Performance in the Limited-Set and Large-Set Trials Separately

a. Limited-Set Trials				
Predictors	Likelihood of Success			
	Odds Ratio	SE	95% CI	<i>p</i>
(Intercept)	0.89	0.08	0.75 – 1.05	.162
Pressure Condition (HPT)	1.30	0.07	1.17 – 1.44	<.001
<i>N</i> _{Subject}	18			
Marginal R ² / Conditional R ²	0.003/ 0.003			
Full vs. null-model	$\chi^2(1) = 24.00, p < .001$			
b. Large-Set Trials				
Predictors	Likelihood of Success			
	Odds Ratio	SE	95% CI	<i>p</i>
(Intercept)	1.97	0.17	1.66 – 2.34	<.001
Pressure Condition (HPT)	1.23	0.07	1.10 – 1.37	<.001
<i>N</i> _{Subject}	18			
Marginal R ² / Conditional R ²	0.002/ 0.002			
Full vs. null-model	$\chi^2(1) = 12.69, p < .001$			

Note. For the categorical predictor of pressure condition the referent category is lower pressure. Significant estimates and their *p*-values are bolded.

Discussion

Overall, we found that monkeys performed significantly worse on limited-set trials than on large-set trials, supporting the hypothesis that capuchins use different memory systems or strategies when faced with a limited or very small set of stimuli versus a large set, in which they can use familiarity rather than working memory or some other effortful system. However, we did not find evidence that monkeys tended to choke under pressure in either limited- or large-set trials. Instead, we found an unexpected tendency to thrive under pressure, such that monkeys were slightly more likely to succeed on the higher-pressure trials than on lower-pressure ones, suggesting that rather than inducing pressure, our training may have served as a motivational cue to achieve a higher reward payout. This differs from our previous results, and we consider this in further detail below.

Our results provide additional evidence, from a new species (tufted capuchin monkeys), that the size of the pool of available stimuli affects the memory systems engaged in DMTS tasks, which has important implications for the interpretation of performance in these tasks, which are prevalent in the literature. Capuchins in the present study showed a drastic drop in performance on the limited-set trials, which mirrors Basile and Hampton's (2013) findings in rhesus macaques, suggesting similar differences in strategy on different stimulus pool sizes across *Platyrrhine* and *Catarrhine* species. Basile and Hampton (2013) also provided evidence that this drop in performance was due to a difference in working memory engagement by including an intermediary task designed to interfere with working memory load during the delay period. While we did not include an intermediary task, as this was not the goal of our study, the fact that our results mirror the dramatic difference that they found suggests that capuchin DMTS tasks with large stimulus pools may also be tapping into familiarity memory rather than working memory. Capuchins' superior performance in the large-set task suggests that they favor a familiarity-memory based strategy when it is possible. Prior work with capuchins which relied on the DMTS as a working memory task may need to be reassessed (i.e., Tavares & Tomaz, 2002, which used 300 stimuli) and future work should both consider this and explore at what size stimulus pool familiarity is no longer a viable strategy for these species to be successful in the task.

Counter to our hypothesis, we did not find evidence of performance failure in the higher-pressure condition. Instead, capuchins were more likely to succeed on the trials that were cued to be higher-pressure, suggesting that rather than acting as stressors, these trials may have been particularly motivating to the capuchins, presumably due to the much higher reward payout (five pellets as compared to one pellet). Further, this pattern held in both the large-set and limited-set trials, despite limited-set trial performance being generally worse. Such results add to a body of literature suggesting that increased reward magnitude does improve performance on a number of cognitive tasks engaging multiple systems, including short-term memory (Basile et al., 2015), temporal interval response behavior (Galtress et al., 2012), and color change detection (Clark et al., 2012). This suggests that pressure does not differentially affect the systems used by the large- and limited-set DMTS trials – that is, simply engaging a working-memory-like system may not be sufficient to induce pressure effects, and that there must other motivational and attentional factors influencing responses to pressure. This fits our previous findings of no effect of pressure overall; however, in that study we also found that individual factors – in that case, hormones – influenced the likelihood of failure under pressure (Sosnowski et al., 2022). There is evidence to support a negative effect of glucocorticoids on memory in humans (for a review, see Lupien et al., 2007), so it is unsurprising that at least some of the variability in pressure responses in that study were associated with individual differences in cortisol. We did not measure hormone levels in the present study, but future work should explore how hormones linked to stress responses or other proximate factors influence in-the-moment failure or success and, in particular, whether they influence different cognitive systems or different levels of pressure differently.

Taken together, these results and those from our previous work suggest that the DMTS task may not be the optimal task to measure pressure effects. Although it seems as though careful curation of the number of possible stimuli can engage working memory in some contexts in the DMTS task, there are other tasks that more directly engage working memory or other executive functions that are prone to interference

from pressure, at least in humans. For instance, tasks in which a series of actions must be carried out in order seem to engage both working memory and procedural memory, and seem prone to interference (for instance, mathematical operations: Beilock, 2008; Beilock & Carr, 2005; or musical performance: Furuya et al., 2021), but analogous tasks have yet to be developed to test pressure effects in non-human animals. However, other species can learn to respond in a certain sequence (for instance, chimpanzees, *Pan troglodytes*: Beran et al., 2004; Inoue & Matsuzawa, 2009; Tomonaga & Matsuzawa, 2000; rhesus macaques: Beran et al., 2004; and tufted capuchins: Beran & Parrish, 2012), so it might be possible to use a similar sequential response task to explore pressure in this context. Indeed, a sequential response task would not only involve working memory even more than the modified DMTS task we used, but would also incorporate an element of procedural memory that might be further sensitive to pressure.

Although future work should certainly explore new tasks in which to study pressure, there is also the possibility that our present results are due to an established cognitive effect related to pressure, but not strictly pressure itself. For instance, the Yerkes-Dodson law, which describes an inverse U-shaped performance curve of affective arousal (Yerkes & Dodson, 1908), would predict poor performance under very low or very high stress conditions. The relatively high performance levels exhibited by our subjects in the higher-pressure trials suggests that it's possible that rather than inducing a high level of negative affect through pressure, we instead induced a more moderate level of arousal, which resulted in better performance than typical on these trials. This could also explain the difference between our previous work and the present study – it might be that upon first presentation in our previous study, the high pressure cues induced a higher level of arousal (resulting in failure under pressure) due to its novelty, but that this prior experience from that study resulted in a more moderate effect in the present work. Another possibility is that our decision to keep high-pressure trials less common than typical trials, while necessary to maintain the higher-pressure nature of the trials, resulted in a von Restorff effect, in which a stimulus or trial that is noticeably different than others in a series is more memorable (von Restorff, 1933). Indeed, this effect has been demonstrated in other non-human primate species (Parker et al., 1998), suggesting that there is a possibility that our subjects might be sensitive to this isolation effect such that they were more attentive during the relatively rare higher-pressure trials. Future work might do well to make higher-pressure trials less distinctive by including background color changes on typical trials that do not indicate a change in reward or consequences.

Overall, our results provide evidence that stimulus pool size is an important consideration when designing DMTS tasks. Capuchins, like rhesus monkeys, appear to use different memory systems depending on whether they can utilize familiarity as a cue when solving a DMTS. However, despite these differences in performance between large- and small-set DMTS tasks, pressure did not affect performance negatively in the present study; in fact, the cued higher-pressure trials resulted in slightly greater success in both large- and limited-set blocks, which suggests that higher rewards likely resulted in higher motivation to succeed. We propose that not only is task demand an important consideration in inducing choking under pressure, but also that motivation as the result of a positive outcome can result in either choking or thriving in high-pressure situations. In particular, the DMTS task may not be the best task to use to study pressure effects, and is therefore not the best choice to induce failure under pressure, at least in the absence of an intermediary task during the delay. Future comparative work in both working memory and pressure effects would do well to consider this when selecting tasks. Further, the development of tasks allowing us to better disentangle working memory from procedural memory will allow us to better understand the effect that pressure has on performance in non-human species, as well as to help us better understand how these cognitive systems may have been conserved among primates.

Acknowledgements

We thank Dr. Sarah Barber, Dr. Marcela Benítez, Dr. Michael Beran, and Dr. Erin Tone for their insightful comments on early drafts of the manuscript. We also thank the CEBUS lab at Georgia State University for their help in running the computerized task. Finally, we thank the Division of Animal

Resources at the Language Research Center for their dedicated and compassionate care for our study subjects.

Author Contributions: MJS designed the study, coded the computer program, conducted data collection and preparation, performed statistical analyses, prepared figures, and wrote the manuscript. SFB designed the study and edited the manuscript.

Funding: During data collection and the writing of this paper, MJS was supported by a Brains and Behavior Fellowship from Georgia State University and a Dissertation Award from the American Psychological Association. SFB was supported by NSF grants SES 1919305 and IBSS 2135621 and the Templeton World Charity Fund (TWCF0471).

Conflict of Interest: The authors declare no conflict of interest.

Data Availability: The data and code that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Basile, B. M., & Hampton, R. R. (2013). Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition*, *126*(3), 391–396.
- Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology: General*, *144*(1), 85.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*(5), 339–343.
- Beilock, S. L., & Carr, T. H. (2005). When high-powered people fail—Working memory and "choking under pressure" in math. *Psychological Science*, *16*(2), 101–105.
- Beilock, S. L., Carr, T. H., MacMahon, C., & Starkes, J. L. (2002). When paying attention becomes counterproductive: Impact of divided versus skill-focused attention on novice and experienced performance of sensorimotor skills. *Journal of Experimental Psychology: Applied*, *8*(1), 6–16.
- Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(6), 983–998.
- Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, *133*(4), 584–600.
- Beran, M. J., & Parrish, A. E. (2012). Sequential responding and planning in capuchin monkeys (*Cebus apella*). *Animal Cognition*, *15*(6), 1085–1094.
- Beran, M. J., Pate, J. L., Washburn, D. A., & Rumbaugh, D. M. (2004). Sequential responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*, *30*(3), 203–216.
- Brady, R. J., Mickelberg, J. M., & Hampton, R. R. (2021). Greater dependence on working memory and restricted familiarity in orangutans compared with rhesus monkeys. *Learning & Memory*, *28*(8), 260–269.
- Clark, A. M., Bouret, S., Young, A. M., & Richmond, B. J. (2012). Intersection of reward and memory in monkey rhinal cortex. *Journal of Neuroscience*, *32*(20), 6869–6877.
- Desrochers, T. M., Burk, D. C., Badre, D., & Sheinberg, D. L. (2016). The monitoring and control of task sequences in human and non-human primates. *Frontiers in Systems Neuroscience*, *9*, 185.
- Dohmen, T. J. (2008). Do professionals choke under pressure? *Journal of Economic Behavior & Organization*, *65*(3–4), 636–653.
- Evans, T. A., Beran, M. J., Chan, B., Klein, E. D., & Menzel, C. R. (2008). An efficient computerized testing method for the capuchin monkey (*Cebus apella*): Adaptation of the LRC-CTS to a socially housed nonhuman primate species. *Behavior Research Methods*, *40*(2), 590–596.

- Furuya, S., Ishimaru, R., & Nagata, N. (2021). Factors of choking under pressure in musicians. *PLOS One*, *16*(1), e0244082.
- Galtress, T., Marshall, A. T., & Kirkpatrick, K. (2012). Motivation and timing: clues for modeling the reward system. *Behavioural Processes*, *90*(1), 142-153.
- Gray, R. (2004). Attending to the execution of a complex sensorimotor skill: Expertise differences, choking, and slumps. *Journal of Experimental Psychology: Applied*, *10*(1), 42.
- Inoue, S., & Matsuzawa, T. (2009). Acquisition and memory of sequence order in young and adult chimpanzees (*Pan troglodytes*). *Animal Cognition*, *12*(1), 59–69.
- Lupien, S. J., Maheu, F., Tu, M., Fiocco, A., & Schramek, T. E. (2007). The effects of stress and stress hormones on human cognition: Implications for the field of brain and cognition. *Brain and Cognition*, *65*(3), 209-237.
- Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., & Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math-anxiety. *Emotion*, *11*(4), 1000–1005.
- Python Software Foundation. (2016). *Python Language Reference, version 3.6*.
- R Development Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Roberts, A. C. (1996). Comparison of cognitive function in human and non-human primates. *Cognitive Brain Research*, *3*(3–4), 319–327.
- Shinners, P. (2011). *PyGame—Python Game Development*. <http://www.pygame.org>
- Smoulder, A. L., Pavlovsky, N. P., Marino, P. J., Degenhart, A. D., McClain, N. T., Batista, A. P., & Chase, S. M. (2021). Monkeys exhibit a paradoxical decrease in performance in high-stakes scenarios. *Proceedings of the National Academy of Sciences*, *118*(35), e2109643118.
- Sosnowski, M. J., Benítez, M. E., & Brosnan, S. F. (2022). Endogenous cortisol correlates with performance under pressure on a working memory task in capuchin monkeys. *Scientific Reports*, *12*(1), 1–10.
- Sosnowski, M. J., & Brosnan, S. F. (2023). Under pressure: The interaction between high-stakes contexts and individual differences in decision-making in humans and non-human species. *Animal Cognition*, *26*(1103–1117).
- Tavares, M. C. H., & Tomaz, C. (2002). Working memory in capuchin monkeys (*Cebus apella*). *Behavioural Brain Research*, *131*(1–2), 131–137.
- Tomonaga, M., & Matsuzawa, T. (2000). Sequential responding to Arabic numerals with wild cards by the chimpanzee (*Pan troglodytes*). *Animal Cognition*, *3*(1), 1–11.
- von Restorff, H. (1933). Über die wirkung von bereichsbildungen im spurenfeld. *Psychologische Forschung*, *18*, 299-342.
- Wittig, J. H., Morgan, B., Masseau, E., & Richmond, B. J. (2016). Humans and monkeys use different strategies to solve the same short-term memory tasks. *Learning & Memory*, *23*(11), 644-647.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, *18*, 459-482.

Supplementary Materials

Table S1

Individual-Level Descriptive Statistics of Proportion Correct In Each Block Type

Subject	Large-Set				Limited-Set			
	LPT		HPT		LPT		HPT	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Albert	0.63	0.11	0.51	0.05	0.43	0.03	0.37	0.04
Bailey	0.56	0.05	0.53	0.01	0.39	0.07	0.35	0.04
Gabe	0.62	0.07	0.65	0.04	0.38	0.01	0.49	0.01
Gambit	0.72	0.06	0.68	0.05	0.61	0.06	0.48	0.09
Gonzo	0.67	0.07	0.82	0.11	0.54	0.07	0.67	0.10
Gretel	0.72	0.12	0.70	0.03	0.53	0.03	0.44	0.08
Griffin	0.73	0.04	0.79	0.04	0.41	0.03	0.56	0.17
Ingrid	0.54	0.07	0.69	0.19	0.33	0.05	0.48	0.08
Irene	0.50	0.02	0.65	0.01	0.44	0.05	0.50	0.15
Ivory	0.79	0.01	0.87	0.10	0.54	0.09	0.64	0.08
Lily	0.57	0.01	0.62	0.03	0.48	0.04	0.56	0.03
Logan	0.67	0.02	0.89	0.01	0.47	0.03	0.78	0.15
Lychee	0.68	0.03	0.74	0.08	0.33	0.02	0.48	0.03
Nala	0.57	0.06	0.67	0.01	0.33	0.02	0.43	0.01
Nkima	0.77	0.14	0.78	0.19	0.62	0.07	0.72	0.00
Paddy	0.69	0.02	0.70	0.03	0.54	0.00	0.57	0.10
Widget	0.78	0.00	0.72	0.00	0.60	0.00	0.53	0.05
Wren	0.64	0.07	0.61	0.19	0.53	0.01	0.57	0.13

Note. (“large-set” or “limited-set”) and condition (lower-pressure: “LPT” or higher-pressure: “HPT”)