AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Comparative Psychology

ISSN: 0735-7036

2021, Vol. 135, No. 4, 545–558
https://doi.org/10.1037/com0000277

The Effects of Positive and Negative Experiences on Subsequent Behavior and Cognitive Performance in Capuchin Monkeys (*Sapajus* [*Cebus*] *apella*)

Mackenzie F. Webster[1, 2] and Sarah F. Brosnan[1, 2, 3]
[1] Language Research Center, Georgia State University
[2] Department of Psychology, Georgia State University
[3] Neuroscience Institute and Center for Behavioral Neuroscience, Georgia State University

Our understanding of animals' affective processing is notably limited compared to the wealth of research on humans, largely due to difficulties in measurement. Moreover, despite a recent increase in the understanding of the interaction between affect and cognition in animals, most research has focused on negative affect, with the result that we continue to know little about the effects of positive affect. In this study, we tested 15 adult capuchin monkeys (*Sapajus* [*Cebus*] *apella*) using a novel methodology that took advantage of capuchins' species-typical behavior to engineer both a positive and negative experience, using the same apparatus to minimize extraneous impacts. Following a positive or negative experience (that presumably induced positive and negative affect, respectively), or a control with no manipulation, we assessed subjects' performance on a cognitive task, a computerized delayed match-to-sample. As predicted, behavior following the negative condition suggested a negative affective state, with increased rates of scratching (commonly used as an indicator of stress in nonhuman primates) compared to both the positive and control conditions. Cognitive performance was also impaired in the negative condition compared to the other two. Contrary to predictions, however, the positive condition did not have a facilitative effect on cognitive performance, but behavioral results indicate that we may not have induced a truly positive affective state. Although we add to evidence that a negative experience can influence subsequent behavior and cognitive performance in nonhuman primates, our work highlights our lack of knowledge about the impact of positive affect, if any, on behavior and cognition.

Keywords: cognition, evolution, frustration, memory, positive affect

Supplemental materials: https://doi.org/10.1037/com0000277.supp

Recent decades have seen an explosion of research on the interplay of emotions, cognitive performance, and behavior (Dolan, 2002; Okon-Singer et al., 2015). In particular, emotion and cognition, once conceptualized as fundamentally different entities, are now considered both interactive and integrated within the human brain (Okon-Singer et al., 2015) and, it appears, some other species' brains as well. For instance, much of the neurological circuitry underlying emotional processing is similar between humans and other species (Panksepp, 2004). New evidence makes clear that animals experience, at minimum, changes in core affect characterized by a spectrum of valence and arousal (Barrett & Bliss-Moreau, 2009; Duncan & Barrett, 2007). Despite this, however, the disparity in our understanding of human emotional processing and that of other animals is considerable. The biggest hurdle to studying affect in animals is measurement. In addition to difficulties with interpretation, most work using noninvasive measures such as scratching and displacement behaviors as measures of negative affect (in primates: Maestripieri et al., 1992; Troisi, 2002) or changes in response to behavioral tests (Paul et al., 2005), lack similar measures for *positive* affect. This is in part because negative affect is often of a higher intensity than positive affect and is much more straightforward to generate. In addition, negative states are typically unwanted and therefore have been a focus of research on how to ameliorate them (i.e., in animal welfare contexts; Boissy et al., 2007). Nonetheless,

This article was published Online First August 26, 2021.

Mackenzie F. Webster https://orcid.org/0000-0002-7795-9282
Sarah F. Brosnan https://orcid.org/0000-0002-5117-6706

The authors wish to thank all the care staff at the Language Research Center for the continual care and concern for the animals they worked with and all of the undergraduate students who helped with the interrater reliability coding, in particular, Jhonatan Saldaña and Caleb Truscott. Sarah F. Brosnan was supported by National Science Foundation Grants SES 1425216 and IBSS 1620391 during this project. Finally, thanks to the monkeys for their participation!

The research described in this paper was Mackenzie Florence Webster's masters thesis project, for which Sarah Frances Brosnan was her supervisor. As such, Mackenzie F. Webster served as lead for conceptualization, analyses, and writing, and developed the methods in conjunction with Sarah F. Brosnan. Sarah F. Brosnan served as the lead for funding and a supporting role for writing and editing.

Correspondence concerning this article should be addressed to Mackenzie F. Webster, Department of Psychology, Georgia State University, 140 Decatur Street, Atlanta, GA 30303, United States. Email: mackenzie.f.webster@gmail.com

we will not fully understand emotional behavior in animals until we also can identify positively valenced affect and understand what causes it.

In most studies of affect in animals, researchers measure behavioral responses or reactions in cognitive tasks after either a naturally occurring or experimentally induced negative experience (i.e., following a fight or an invasive veterinary procedure, respectively) that is presumably generative of a negative affective state or following the presentation of negative valence visual stimuli. For instance, research with rhesus monkeys (*Macaca mulatta*) demonstrated an impaired ability to locate hidden food after a delay when experiencing acute noise stress (Arnsten & Goldman-Rakic, 1998). Although most of this research on the interaction of affect and cognition has focused on primate species, several studies have also shown that acute stress impairs both spatial and recognition memory in rodents (for review, see Cazakoff et al., 2010); thus the findings may be expandable to a broader comparative network. Rhesus monkeys and Japanese macaques (*Macaca fuscata*) show attentional biases to both threatening social (Lacreuse et al., 2013) and nonsocial stimuli (Shibasaki & Kawai, 2009). An emotional Stroop experiment with chimpanzees (*Pan troglodytes*) found that a negatively valenced photograph (of a veterinarian) interfered with performance and increased latency on a color discrimination task (Allritz et al., 2016). Baboon (*Papio papio*) reaction times (RT) on a well-trained cognitive task were influenced by the subjects' mood (inferred from the affective valence of spontaneously occurring behavior), increasing when experiencing a negative mood compared to positive or neutral (Marzouki et al., 2014). Finally, rhesus monkeys showed social attentional effects, in the form of faster gaze disengagement and less overall gaze directed toward an aggressive face, following a veterinary procedure presumed to induce anxiety and stress (Bethell et al., 2012). These results all point to an important interaction between affect and cognition.

What is strikingly missing from the literature, however, is the impact of *positive* affect or experiences on cognition in animals. Of the studies discussed, only two looked at the effects of both positive and negative affect on cognition. Baboons did not demonstrate a decrease in RT when in a positive mood (as assumed to occur following the display of behaviors associated with a positive valence; Marzouki et al., 2014). Rhesus monkeys, on the other hand, showed sustained attention to a threatening face following enrichment compared to that following a stressful veterinary procedure (Bethell et al., 2012). Although these latter results indicate a possible effect of positive emotional states on primate social attention, the authors refer to this condition as a "neutral (or potentially positive)" condition, observing fewer behavioral indicators of stress than after the presumably negative condition. Nonetheless, the absence of negative affect is not sufficient to indicate positive affect. Therefore, unless we wish to argue that positive affect does not interact with cognition as negative affect does, and human responses suggest no theoretical reason to assume that this would be the case, we are clearly not properly measuring or generating positive affect in these studies. Although recent years have seen a surge of research aimed at discovering ways to measure and generate positive affect, primarily in the animal welfare literature, little attention has been given to the interaction of positive affect and cognition.

Indeed, there is not widespread consensus on what might universally generate positive affect across nonhuman species, either in natural contexts or experimentally. In studies looking for interactions between affect and cognition, researchers used spontaneous behaviors hypothesized to be associated with a positive valence (Marzouki et al., 2014) and the provision of enrichment (Bethell et al., 2012) to identify (possible) positive affect. However, these are quite different procedures, both from each other and from the procedures used to induce negative affect. Other researchers studying positive affect in animals have similarly used a wide range of procedures, such as the delivery of food reward, environmental enrichment, video content, and even the absence of negative outcomes, to generate what they assume is positive affect, or they inferred it after naturally occurring presumptive positive experiences, such as receiving grooming from a conspecific. One promising possibility coming out of the animal welfare literature is the idea that problem-solving opportunities may act as a form of enrichment (Meehan & Mench, 2007). Indeed, research with cattle and dogs has found some behavioral evidence that positive affect is generated in response to an animal's own achievement in obtaining a reward through performing an operant task (Hagen & Broom, 2004; McGowan et al., 2014).

Part of the problem is that this assortment of procedures used to induce affect—either positive or negative—likely generates experiences that differ in both magnitude and type, making it difficult to determine whether a given experience is sufficiently strong to generate an effect. In addition, the type and intensity of the affect-inducing experiences may differ for positive and negative experiences even within the same study, introducing confounds. For instance, in Marzouki and colleagues' (2014) work, behaviors used to infer a positive mood were primarily social behaviors (such as social grooming and play), whereas the behaviors used to infer a negative mood were primarily nonsocial (such as stereotypies and body shaking), confounding mood induction and social context. Further, the intensity often varies between the conditions, confounding valence and intensity. For instance, the negative condition may consist of a strongly negative experience, such as an invasive veterinary procedure, whereas the positive is a relatively standard enrichment experience (Bethell et al., 2012). Although these studies provide important insight into the effects of specific positive and negative experiences on behavior and cognition, it is difficult to isolate the effects of the changes in affect specifically using such different procedures to induce positive and negative affect. Of course, a final challenge is that because we do not have a method of measuring positive affect, even if we induce it, we may not know that we did so.

Thus, the purpose of the current study was twofold: to test the effects of different experiences that we predicted would induce positive or negative affect in the monkeys on subsequent performance on a cognitive task and to look for reliable behavioral measures of affect. In particular, we looked for behavioral measures that may be associated with a positive experience (and therefore potentially of positive affect). To do this, we created a new paradigm that we hypothesized (1) would engender negative as well as positive experiences in our subjects, using as similar a procedure as possible to limit alternative explanations and 2) reflects natural experiences in the subjects' lives that would be akin to typical daily fluctuations in a healthy organism's life (of similar intensity and not extreme).

We tested brown capuchin monkeys (*Sapajus* [*Cebus*] *apella*), a New World primate species with a large brain-to-body ratio, equivalent to that of great ape species (Rilling & Insel, 1999). They have been widely studied in the laboratory and wild and show sophisticated cognitive and social behaviors (Amici et al., 2012; Fragaszy et al., 2004; Seed & Tomasello, 2010). In designing our task, we aimed to create a species-appropriate manipulation that reflected natural experiences in our subjects' lives. This was important to minimize training, increase ecological validity, and keep the affective experience within the range of normal daily fluctuations. To do so, we took advantage of the fact that capuchins are extractive foragers, extricating food from difficult to obtain sources, such as inside hard-shelled nuts (Fragaszy et al., 2004). To mimic this, we used a food retrieval puzzle with which we could artificially create a negative experience in which food was unobtainable from a previously reliable source (such as an impossible to crack nut), and a positive experience in which a highly-preferred food could be retrieved.

To measure how our procedure influenced performance on a cognitive task, the experimental manipulation was immediately followed by a computerized delayed match-to-sample (DMTS) task to measure cognitive performance in the monkeys. The DMTS task has been widely used to study working memory across a range of animal species (for a review of cross-species DMTS data, see Lind et al., 2015) and is a task that our capuchin monkeys were already familiar with. We predicted that performance on the DMTS task would be impaired in the negative condition, in keeping with prior research showing a negative effect of stress on cognitive performance (Arnsten & Goldman-Rakic, 1998; Cazakoff et al., 2010). Because the literature on positive affect in animals in general is sparse, our predictions for the effects of the positive experience on cognitive performance were based entirely on the human literature. Yang et al. (2013) found improved working memory in humans after receiving a small gift, and other research has consistently found evidence that positive feelings affect cognitive processing (for a review, see Ashby et al., 1999). Based upon this, we predicted an enhancing effect of the positive experience on cognitive performance.

Our second goal was to find a behavioral measure (or measures) associated with the positive experience that indicated positive affect and could be used to assess affect in group settings and noninvasive contexts. Because we had no data from the literature on which to base hypotheses, we used an exploratory approach, collecting data on all behaviors exhibited by the subjects, whether or not they are discussed in the literature on animal affective experience. Because this was by nature an exploratory study, we did not have any direct predictions for what behaviors might increase following the positive experience. Consistent with the literature, we predicted that rates of displacement behaviors, such as scratching, self-touching and urine washing, as well as stereotypic behaviors would increase after subjects had a negative experience that presumably led to negative affect (Garner, 2005; Maestripieri et al., 1992; Troisi, 2002); and we expected these results to serve as a manipulation check. Moreover, although a major goal of the behavioral measurement aspect of the study was to search for behavioral indicators of positive affect, this also served as a manipulation check to help determine if a positive affective state was induced if we failed to see changes in performance on the DMTS task.

Method

Subjects

Subjects were 15 adult (seven male) brown capuchin monkeys (*Cebus [Sapajus] apella*) housed at Georgia State University's Language Research Center (LRC). Subjects were all socially housed in one of three separate mixed-sex social groups composed of five, six, and 10 individuals (Groups 1, 2, and 3, respectively). As is true in the wild, most of the subjects were born into and grew up in their social group. All subjects were mother-reared. Subjects' housing enclosures included both an indoor room and a large outdoor area, both of which had enrichment toys and climbing structures for natural movement and activity. Subjects received primate chow, fruits, and vegetables daily and supplemental enrichment foods most days. Subjects were never food deprived, and water was available ad libitum, including during testing. All procedures used in the research are in accordance with the Guidelines for the Use of Animals in Research of the United States and have been approved by the Institutional Animal Care and Use Committees of Georgia State University (IACUC protocol A13020).

All subject participation in this study was voluntary. This is always the procedure at the LRC, but is particularly relevant for a study of affect, where any stress from the process of testing would interfere with results. As with any study at the LRC, subjects expressed their willingness to participate by voluntarily separating into individual test boxes connected to their indoor home enclosure on the morning of testing days. Subjects received either one peanut (Groups 1 and 2) or one pecan (Group 3) upon entering the individual test box. Subjects who chose not to participate did not receive any consequences, except not being able to participate in the study (i.e., they were never restricted from food, treats, groupmates, or outdoor access for choosing not to participate). Subjects at this facility readily separate into these individual test boxes daily, with no knowledge of which task they will be interacting with (tasks typically vary from day to day), making self-selection unlikely. Of the 21 subjects available, only 15 were used for the current study, as three chose not to voluntarily separate on a regular basis, one was excluded due to impaired visual and motor abilities as a result of old age, and two monkeys failed to pass the training criterion for the delayed match-to-sample procedure (see the following sections).

General Procedure

For this study, subjects participated in the experimental manipulations, followed by a computerized cognitive task (DMTS). Subjects were given 18 sessions, six positive, six negative, and six control sessions, presented in a pseudorandomized order so that half the subjects ($n = 8$) experienced the negative experience first and the other half ($n = 7$) experienced the positive experience first. All subjects were presented with the six control sessions in between the positive and negative sessions. Subjects were never presented with more than one session per day. The same apparatus was used for both the positive and negative manipulations to minimize any extraneous variables that differed between the two conditions. Following experimental manipulations (or the control), subjects were immediately placed onto the computerized test. All sessions were videotaped for later behavioral coding. Subjects were first trained on the puzzle apparatus used for the experimental manipulations, then testing commenced; when it became clear during initial testing that they needed additional training on the computerized DMTS task, they received the additional training at that point (see the following sections for details).

Apparatus

The apparatus (Figure 1) was an opaque tube that attached diagonally to the outside of the subjects' individual test boxes, hung at an angle such that a food reward placed in the top would roll down it into a tray from which the subjects could retrieve it. Three slides, painted blue, were placed perpendicularly at equal lengths down the tube such that when closed, they would stop the progression of rolling food. To retrieve a food reward placed at the top, subjects had to lift each slide in successive order to move the food reward down into the retrieval tray (Figure 1b). Food rewards in this study were small pieces of colorful cereal (Cap'n Crunch OOPS! All Berries Cereal, a preferred food).

Puzzle Apparatus Training

Subjects were first presented with the puzzle apparatus for training/familiarization sessions before testing. Each session consisted of 10 trials. At the start of a trial, all three slides (located next to the arrows in Figure 1a) were closed. Trials began when the tube was baited with the food reward, and the subjects had 1 min to retrieve it. If the subject was unable to complete the trial within the allotted time, the experimenter lifted each slide herself, allowing the subject to watch the reward fall into the retrieval tray. Subjects were then allowed access to the reward and incurred a 10-s time-out before the next trial. Subjects passed the training phase when they successfully retrieved the reward (without the experimenter's assistance) on at least eight out of 10 trials across two consecutive sessions. All subjects successfully passed training within three sessions (two sessions were the minimum possible number of sessions to pass criterion), indicating that the task was fairly intuitive to the animals.

Puzzle Apparatus Testing

Testing was identical to training with a few small differences. During testing, subjects had 30 s (instead of 1 min) to retrieve the food reward. If the subject had not been successful within the 30-s period, the trial would have timed out and there would have been an additional 15-s time-out before the next trial began (although this never occurred during testing sessions). If a trial was successful, subjects experienced a shorter, 10-s intertrial interval (ITI) before the next trial began. Different manipulations were used to generate the positive and negative experiences.

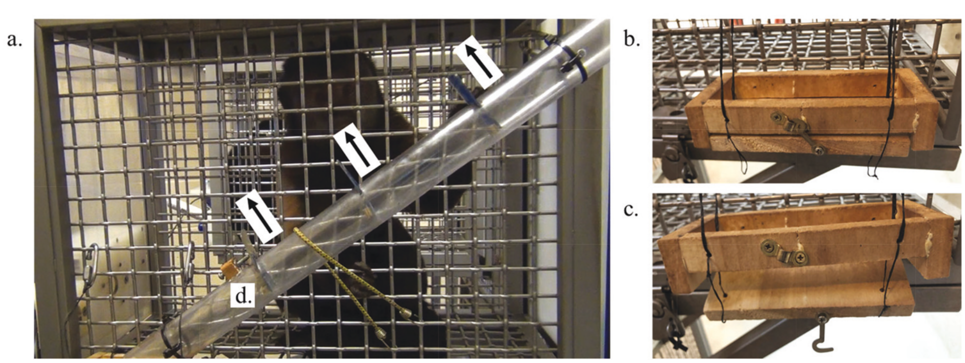
Positive Experience

To create a positive experience for the monkeys, subjects were presented with the familiar puzzle apparatus (see Video S1 in the online supplemental materials). Subjects completed six consecutive trials with positive reinforcement in the form of the preferred food reward. Consistently receiving a desired food reward was assumed to be a positive experience. It has also been demonstrated that object manipulation tasks are mentally enriching to primates (Celli et al., 2003), and recent research in the animal welfare literature has begun to indicate a connection between problem-solving and positive affect (Hagen & Broom, 2004; McGowan et al., 2014; McGowan et al., 2010; Meehan & Mench, 2007), so this was predicted to further influence the positivity of the experience.

Negative Experience

To generate a negative experience, the subjects were presented with the same puzzle task, but successful completion of the task (and therefore access to the food reward) was impossible on some trials. Subjects completed 10 consecutive trials during each negative session. During four of those trials, however, the food reward was unobtainable, a presumably frustrating experience

Figure 1
Puzzle Apparatus Used for Experimental Manipulations



Note. Arrows in (a) indicate the direction the three slides lifted in order for the food reward to pass down the tube, traveling from the top right in the picture to the bottom left, where the reward would fall into the retrieval tray (b) and be collected (see text for detail). (c) depicts how the tray opened in the "trap-door" sessions. (d) indicates the location of the screw that would allow the third slide to be jammed in the "jammed-slide" sessions. See the online article for the color version of this figure.

for the subjects. As subjects were unable to acquire the reward in 4/10 trials, the negative sessions resulted in a total of six food rewards, the same as in the positive condition, to control for satiation. The impossible trials occurred in a pseudorandomized order within the 10-trial sessions, with no more than two impossible trials in a row. Only four impossible trials were included in the negative condition to frustrate the subjects without causing them to cease participating.

There were two different types of impossible trials: jammed-slide trials and trap-door trials. Negative sessions consisted of *either* jammed-slide or trap-door trials (three sessions of each type presented in a randomized order). During jammed-slide trials, the apparatus was experimentally manipulated by turning a screw (Figure 1d) so that the third and final slide was prevented from lifting high enough to allow the food reward to pass underneath. Trials lasted 30 s, after which the slide was unjammed and the food reward was removed by the experimenter (see Video S2 in the online supplemental materials). In trap-door trials, a latch in the bottom of the retrieval tray was unhooked so that after successful lifting of the third slide, the food reward fell through the retrieval tray to the floor, where it was unreachable (Figure 1c; see Video S3 in the online supplemental materials). During the negative sessions, before every trial (impossible or not), the experimenter manipulated the screw (jammed-slide) or the retrieval tray latch (trap-door) so that subjects were not cued to when the impossible trials were set to occur (by making the same movements and sounds before both types of trials). We cannot be certain the subjects were completely unaware of which trials were the "bad" ones, however, in every case all subjects attempted to solve each trial (impossible or not). The two different types of negative experience (trap-door and jammed-slide) were included to ensure the generation of a negative experience, in case one method may have been more effective than another, but there were no prior predictions on which would be more successful. Subsequent analysis revealed no significant differences between the two types of negative experiences on performance, so all further analyses grouped the two manipulations into a single negative condition.

Control

We included a control condition to create as "neutral" an experience as possible. In this condition, subjects were placed directly onto the cognitive computer task after entering their individual test boxes. Subjects were not given access to the puzzle task, as a manipulation of any kind with the tube without receiving a reward might be interpreted as a negative experience. The subjects were also not given any food reward prior to testing, as receiving a "free" food reward might skew the interpretation toward the positive. Although this meant we did not control for satiation between the control condition and the two experimental conditions, subjects were only able to retrieve six small pieces of cereal from the puzzle apparatus in the other conditions, and we felt this small difference would have less impact on behavior or cognition than the alternatives.

Cognitive Test (DMTS)

Immediately following the experimental manipulations, the tube was removed and subjects were presented with the computerized test of cognitive performance (DMTS task) within the same individual test box as the manipulations. Each subject had a personal computer with a 17-in. monitor, a joystick that moved a cursor on the screen, and a pellet dispenser that distributed small 45-mg banana-flavored pellets to the subjects as a reward for a correct response. Personal computers were stationed approximately 30 cm in front of a Plexiglas window on the individual test box, with their personal joystick placed in the box with them. Subjects at the LRC have had years of experience testing and were therefore extremely familiar with the computerized system. They also had extensive prior experience with the DMTS (Beran et al., 2008; 2012; Beran & Smith, 2011; Evans et al., 2008).

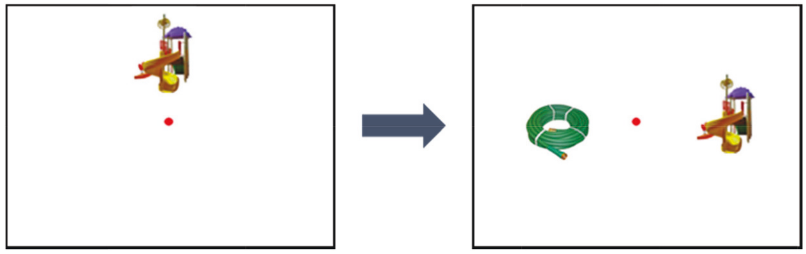
DMTS Training

Subjects were not originally expected to require training on the cognitive task, as all the monkeys had passed a computer training battery program that included a DMTS at some point before the onset of the study. However, for some subjects it had been some time since they had experienced a DMTS task, and in the initial presentation of this study, they were presented with complex novel stimuli. An analysis of the data from the subjects' first three or four sessions (six monkeys experienced three sessions, the other nine experienced four) revealed that the subjects were performing below 80% accuracy. We did not want learning to be a factor in the study, so this initial testing was immediately stopped, and the subjects were placed on training to get them to an 80% accuracy criterion. None of these data were used in the analysis.

Training consisted of presentation of the computerized task for 1-hr sessions, during which subjects could do as many or as few trials as they chose. At the onset of a trial, an image appeared centered near the top of the subject's screen, along with their cursor, also in the center. The stimuli used in training were simple clipart images that the subjects had previous experience with from their original DMTS training (in contrast to the complex stimuli used in the initial testing that may have impaired performance). Subjects could move their cursor up toward the image, and upon contact with the border of the image, the picture disappeared. This was then followed by a delay of 1, 2, 3 or 5 s (in randomized order, with five trials of each time delay presented in 20-trial blocks). After the delay, the target image and another picture simultaneously appeared, randomized between the right and left sides of the screen (Figure 2). If the subject moved the cursor to the image that matched the original sample, they heard a familiar chime indicating a correct response and received a small 45-mg banana flavored pellet. This was followed by a 2-s ITI before the next trial began. If the subjects moved their cursor over the incorrect image, they heard a buzzer indicating an incorrect response and received no reward and a 5-s timeout before the next trial. Regardless of accuracy, after every 20 trials, there was a 1-min break in the task.

Our delay periods (1, 2, 3 and 5 s) for the DMTS were chosen based on a previous study with capuchin monkeys that indicated that 3-s delays affect both subject response time and accuracy as compared to shorter time delays (Truppa et al., 2014). The 5-s delay was also included to see if there were any effects of the experiences on a more challenging task, without increasing the duration so much so that performance decreased below 80% (as is evidenced after 8-s delays, although performance often remains above chance [50%] for much longer delays; Tavares & Tomaz, 2002). Although we did not predict any differences in effect across

Figure 2
Example of the Delayed Match-To-Sample Task Screen



Note. Examples of the BOSS photos used as stimuli during delayed match-to-sample training and testing. See the online article for the color version of this figure.

the different delay periods, we included a range so that we could determine whether any effects were linked to task difficulty (and intentionally kept our range within that which capuchins are known to easily solve to avoid additional frustration).

Subjects reached training criterion once they demonstrated 80% or better accuracy during two consecutive sessions. They were then placed on one or, if needed, two probe sessions, where subjects again experienced a 1-hr session, but this time with novel stimuli. Instead of the complicated novel stimuli used in the original testing, for probe trials, we used a bank of 238 images from the Bank of Standardized Stimuli (examples in Figure 2), a large normative photo database (Brodeur et al., 2014). If subjects successfully maintained 80% or higher accuracy on the first or second probe session (all subjects did), then they passed to the testing phase. Overall, it took the subjects a mean of 11 training sessions (σ = 12.4) to reach the training criterion (ranging from two to 40 sessions between individuals). The fact that they took so long to reach criterion supports our decision not to count the original data.

DMTS Testing

Testing followed the same procedure as training but used a bank of 450 *different* BOSS photos than were used during the probe sessions. Test sessions were 30 min in duration. Although the literature on humans indicates that the temporal retention of most types of experimentally manipulated affect may be less than 30 min (Frost & Green, 1982; Gomez et al., 2009), other research specifically relating to anxiety indicates that the duration of an anxious mood could last longer (Kuhlmann et al., 2005; Schoofs et al., 2008). Although the effects of our manipulation were expected to be fairly short-lived, the 30-min testing session was selected to ensure that any longer lasting effects were caught, as well as to maintain consistency with typical testing procedures at the facility. Digital time stamps for each trial were recorded, which allowed us to go back and look at accuracy over shorter time blocks within the session. Ultimately, we analyzed the session in three 10-min sections to look for effects of time.

Behavioral Analysis

To capture possibly unknown behavioral indicators of positive affect, every behavior observed during the DMTS test sessions was recorded. An ethogram was created from a list of previously known capuchin behaviors, and any novel behaviors observed during a preliminary review of the videos were added (no new behaviors were observed after one session of each subject was reviewed). The final ethogram included a total of 19 behaviors (Appendix Table A1).

Of the 19 behaviors recorded, we eliminated four that were observed in fewer than 10 of the 15 subjects. In two cases, several behaviors were combined because there were strong theoretical grounds to do so (this also increased power). Specifically, "licking the cage," "picking at the cage," and "wiping the cage" were all grouped together into "cage-directed behaviors." "Pacing," "head-twirling," and "rub hands" behaviors were grouped together into a "stereotypic behavior" category, as these are all stereotypical, repetitive behaviors that often manifest in captive primate populations (Garner, 2005; Garner et al., 2003; Pomerantz et al., 2012). These stereotypic behaviors alone would have each been eliminated from analysis based on the criterion that at least 10 subjects display the behavior, but as a category, this involved 10 or more individuals. Therefore, for the ultimate analysis we looked at nine individual behaviors plus these two behavioral categories.

In all, 20% of the videos were double-coded by a second rater blind to the study's hypotheses to establish IRR. IRR was calculated for each behavior using correlations between the numbers of behaviors indicated by the main experimenter per session compared to the IRR coder. IRR is reported for each behavior in the Results section. Using this approach, more objective and commonly observed behaviors had very high IRR. On the other hand, some infrequent behaviors had very low IRR. We chose not to recode those behaviors, as it seemed to us to be important that these behaviors were so much less reliable, despite being coded by the same experimenters using the same approach, which suggests that even if we had managed to get higher IRR, they might not be appropriate behaviors to use. Nonetheless, to be complete we report on all behaviors, but do not further consider those with low IRR.

Data Analysis

All statistical analyses were run in R (R Core Team, 2015). Model analyses were conducted using the *lme4* package (Bates et al., 2015). All model comparisons were compared using the Akaike information criterion to determine the best-fit models; *p* values were determined via likelihood ratio tests comparing the full model with the fixed effects against a null model with just the random effects. Behaviors were measured as counts of every occurrence of the behavior (or behavioral category) of interest within a session, which began immediately following the experimental manipulations at the onset of the 30-min DMTS task. Sessions were divided up by time into three sections (0–10 min, 10–20 min, and 20–30 min to examine any possible effects of time within a session. Accuracy was measured as the proportion of correct responses to the total number of trials completed. Visual inspection of residual plots did not reveal any obvious deviations from homoscedasticity or normality.

Results from the different delay periods indicated that performance on 1- and 2-s delay trials did not differ from one another, but performance was impaired by 3-s delays and further compromised by 5-s delays. There was not an interaction between delay and condition on their impact on performance, so for all subsequent analyses, data from all delays were combined to provide maximum power to the overall analysis.

First, to assess whether behaviors differed across the conditions, we ran a series of generalized linear mixed models with Poisson distributions for each behavior. We constructed separate models with counts of each behavior (or behavioral category) as the dependent variables (DVs), condition (control, positive and negative) as a fixed effect, and Subject ID entered as a random effect. Analysis excluded three cases with missing data. We compared the full model to the null model (which included only the random effect) using a likelihood ratio test.

Second, to determine if condition influenced overall levels of participation in the DMTS task, we ran a generalized linear mixed model with trial number as the DV. We included condition and time as fixed effects and Subject ID as a random effect.

Third, we examined the effects of both condition and time on accuracy. To assess whether condition and time influenced accuracy, and whether there was an interaction, we constructed a model comparison analysis comparing four linear mixed models (LMM) with accuracy as the DV, two models with each condition or time as fixed effects alone, a combined model with both factors, and a model with an interaction effect. All controlled for Subject ID as a random effect.

To determine whether the behaviors influenced by condition were also influencing DMTS task performance, we ran a series of LMMs with accuracy as the DV, each behavior as a fixed effect, and Subject ID as a random effect. The behaviors that significantly predicted accuracy were then further analyzed with another series of LMMs, comparing three models for each behavior's effects on accuracy (DV); a model that included just the behavior, one with both the behavior and condition as fixed effects, and one with an interaction effect (all with Subject ID as a random effect).

Finally, to determine the overall best model for predicting cognitive performance, a model comparison was conducted comparing including each of the fixed effects (condition, time, and the behaviors that significantly influenced performance) with and without each other, as well as models with any significant interactions.

Results

Behavior

Of the 11 behaviors (or behavioral categories) analyzed, seven significantly differed between conditions (Figure 3), scratching: $\chi^2(2) = 12.63$, $p = .002$ [IRR; $r = .95$]; self-licking: $\chi^2(2) = 15.17$, $p < .001$ [IRR; $r = .79$]; threatening: $\chi^2(2) = 18.98$, $p < .001$[IRR; $r = .37$]; self-touching: $\chi^2(2) = 12.90$, $p = .002$ [IRR; $r = .61$]; playing with pellet: $\chi^2(2) = 23.83$, $p < .001$ [IRR; $r = .81$]; cage-directed behavior: $\chi^2(2) = 82.62$, $p < .001$ [IRR; $r = .95$]; stereotypic behavior: $\chi^2(2) = 161.95$, $p < .001$ [IRR; $r = .87$]).

Most of our effects were due to differences between the control and both of the two experimental conditions (positive and negative), suggesting that any intervention changed behavior, regardless of its intended valence. Specifically, we found that compared to the control condition, in both the positive and negative conditions there was significantly more stereotypic behavior (Figure 3a; negative: $\beta = .26$, $z = 8.66$, $p < .001$; positive: $\beta = .36$, $z = 12.37$, $p < .001$) and self-touching (Figure 3e; negative: $\beta = .26$, $z = 3.50$, $p = .001$; positive: $\beta = .18$, $z = 2.49$, $p = .034$), and significantly less cage-directed behaviors (Figure 3c; negative: $\beta = -.19$, $z = -6.51$, $p < .001$; positive: $\beta = -.26$, $z = -8.65$, $p < .001$) and playing with the pellet (Figure 3d; negative: $\beta = -.14$, $z = -3.32$, $p < .001$; positive: $\beta = -.20$, $z = -4.73$, $p < .001$).

As predicted, however, there were some changes in the negative experimental condition as compared to both the positive and control conditions, and in both cases, these were behaviors that are consistent with our prediction that the experience generated a negative affective state. Specifically, compared to the negative condition, in both the positive and control conditions, there was significantly less scratching (Figure 3b; positive: $\beta = -.08$, $z = -3.04$, $p = .007$; control: $\beta = -.09$, $z = -3.10$, $p = .005$) and threatening behavior (Figure 3f; positive: $\beta = -.51$, $z = -2.55$, $p = .028$; control: $\beta = -.97$, $z = -4.14$, $p < .001$), although the IRR for threatening behavior was too low to justify further consideration ($r = .37$). There was also, surprisingly, more stereotypic behavior in the positive condition than the negative (Figure 3a; $\beta = .10$, $z = 3.69$, $p < .001$); however, there was still significantly more stereotypic behavior in the negative condition than the control ($\beta = -.26$, $z = -8.66$, $p < .001$).

The only behavior significantly different in just the positive condition compared to the other two was self-licking behavior (Figure 3g), of which there was significantly more in the positive condition than either the negative condition ($\beta = .17$, $z = 3.28$, $p = .003$) or the control ($\beta = .18$, $z = 3.42$, $p = .002$).

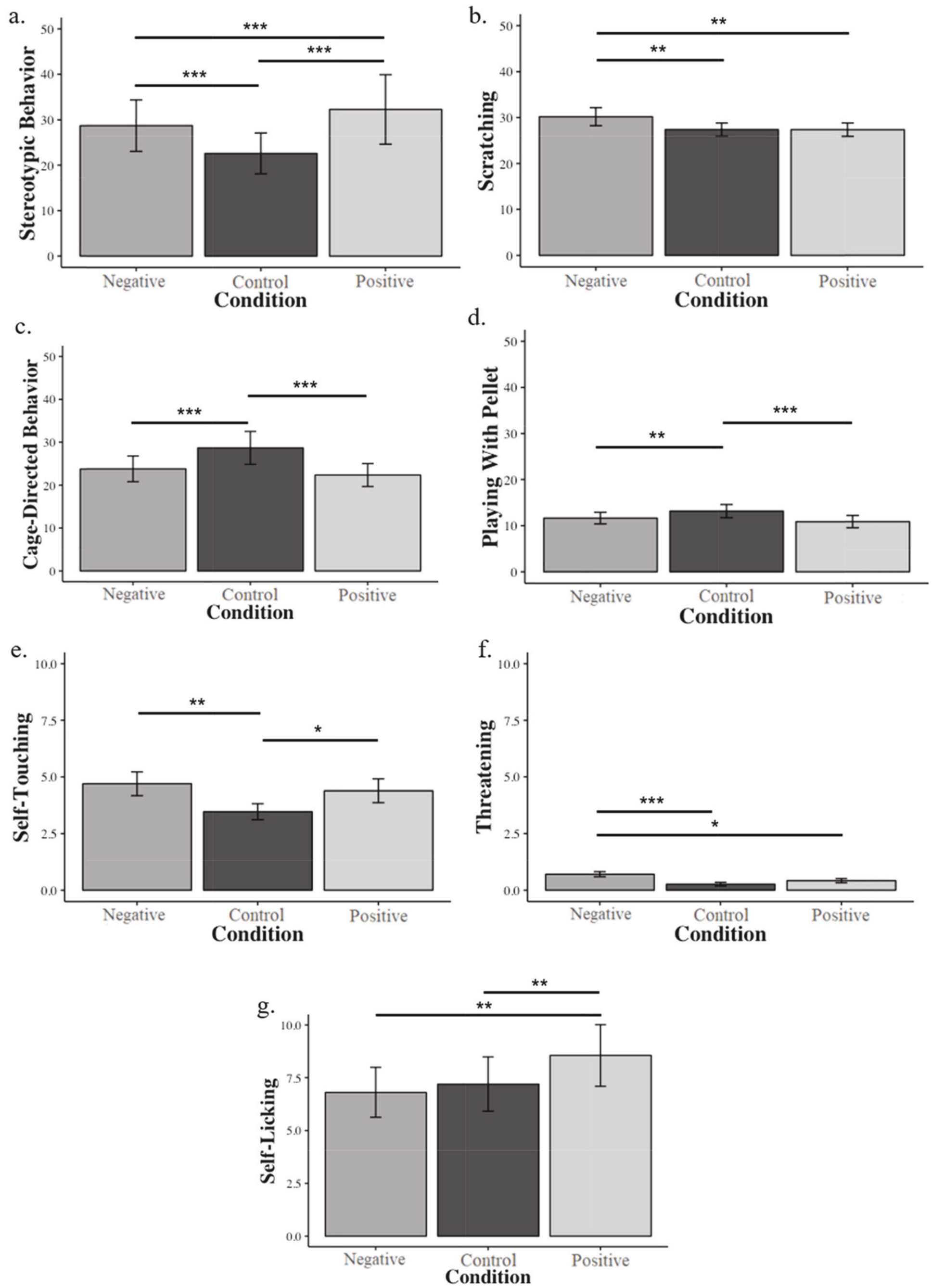
Cognitive Task (DMTS) Performance

The best-fit model for predicting the number of completed trials (trial number) included both condition and time as fixed effects, better-explaining trial number than models with either factor alone or a null model, however there was no interaction effect between condition and time (Table 1). On average, subjects completed 99.4 (± 2.97) trials in the positive condition, significantly fewer than the number of trials completed in the negative condition (103.0 ± 2.91 trials; $b = -.05$, $z = -3.38$, $p = .002$) or in the control (106.0 ± 3.31 trials; $\beta = -.07$, $z = -4.72$, $p < .001$; Figure 4). Regardless of condition, subjects completed the most trials during the first 10 minutes of sessions (36.9 ± .54), significantly fewer during the middle third of the session (34.2 ± .72 trials; $\beta = -.08$, $z = -5.33$, $p < .001$), and the fewest number of trials during the last 10 minutes (31.3 ± .80 trials; $\beta = -.09$, $z = -5.98$, $p < .001$), although the decrease is modest, suggesting that they are fairly motivated for the entire 30 min.

Overall, subjects performed quite well on the DMTS task, with a composite average of 83.4% (± .40) correct. This was as expected, as the criterion to participate in the study was 80% or greater

Figure 3

Results for the Average Occurrence of Behavior Across Condition



Note. Figures depict average behavioral occurrences between conditions for stereotypic behavior (a), scratching (b), cage-directed behavior (c), playing with the pellet (d), self-touching (e), threatening (f), and self-licking (g). Graphs depict mean aggregated summary results and do not control for subject differences, whereas the p values and bars representing significant differences come from generalized linear mixed models that control for Subject ID as a random effect. Y axis change between graphs A–D (ymax = 50) and E–G (ymax = 10). Note that the interrater reliability for threatening behavior was too low to be further considered in our analyses. Error bars reflect ± 1 *SE*.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 1
Comparison of the Models Used to Predict Trial Number

Model	*df*	AIC	χ^2	*p*
Null		7,569.9		
Condition only	2	7,550.2	23.63	<.001*
time only	2	7,445.9	127.97	<.001*
Condition + Time	2	7,426.3	23.63	<.001*
Condition × Time	4	7,426.1	8.12	.087

Note. AIC = Akaike information criterion.
* $p < .05$.

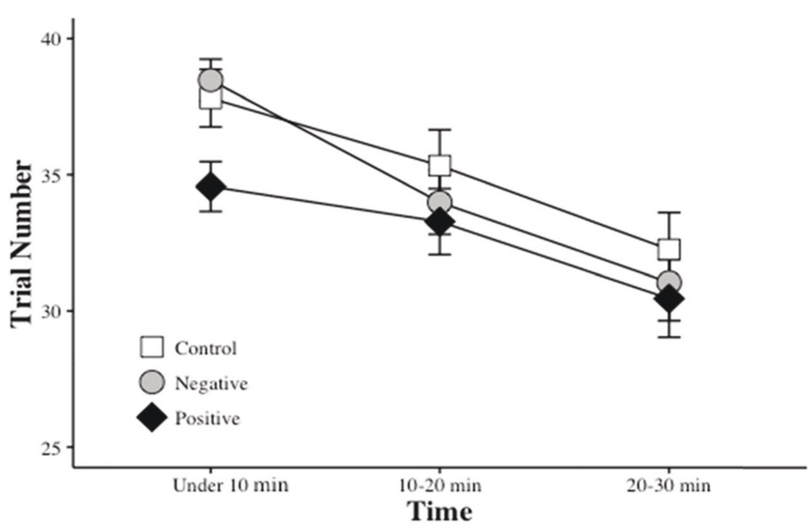
accuracy. There was still, however, a significant effect of condition on cognitive performance, $\chi^2(2) = 12.43$, $p = .002$, with higher accuracy in the positive condition (84.1 ± .64%, $\beta = .02$, $t = 2.96$, $p = .003$) and control (84.5 ± .67%, $\beta = .03$, $t = 3.09$, $p = .004$) compared to the negative condition (81.5 ± .75%). There was no significant difference in performance between the positive condition and the control, and there was no significant effect of time on accuracy or an interaction between time and condition (Figure 5).

Of the seven behaviors significantly affected by the experimental manipulation, three also significantly correlated with performance on the DMTS task. Higher rates of playing with the pellet ($\beta = .001$, $t = 3.66$, $p = .003$, Figure 6b) and self-licking behavior ($\beta = .002$, $t = 3.84$, $p < .001$, Figure 6c) related to higher levels of accuracy on the task. Contrasting this, increased rates of scratching correlated with decreased performance on the DMTS task ($\beta = -.001$, $t = -2.60$, $p = .010$, Figure 6a), which was consistent with our findings of increased scratching and impaired performance in the negative condition. Increased rates of stereotypic behavior also related to decreased performance; however, the effect was not statistically significant ($p = .055$). This may, however, have been related to the lack of enhanced performance in the positive condition, as that is the condition in which we saw the most stereotypic behavior. Although we cannot reliably use threatening behavior, we do note that there was not a significant correlation of accuracy with threatening behavior, which may be the result of both low baseline occurrence rates and a low IRR, again suggesting that it may not be a good measure of affective state in capuchins.

Because both behavior and condition significantly predicted cognitive performance, we ran model comparisons to determine whether condition and the behaviors were independently impacting performance or whether the effect was simply the result of an interaction. Although each model predicting accuracy with behavior alone as a factor was significantly better than a null model, and every model was improved by the addition of condition as a fixed effect, none of the models showed a significant interaction (Table 2).

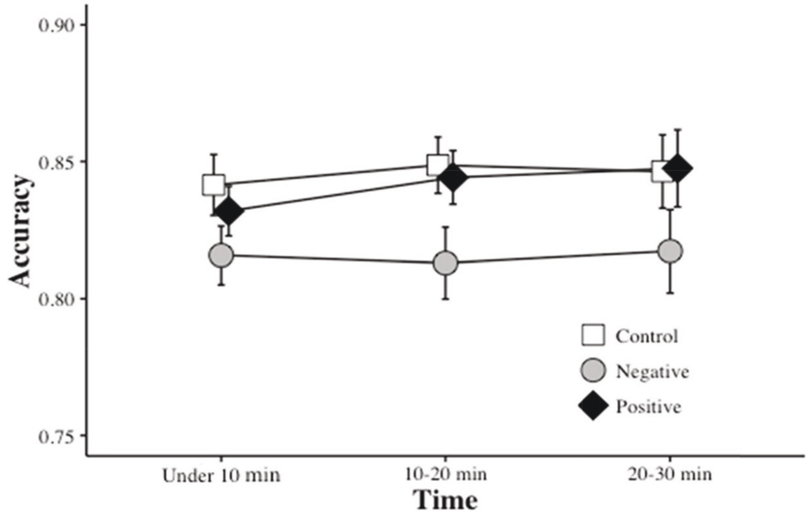
The overall best-fit model for predicting accuracy on the DMTS task included condition and the behaviors scratching, self-licking, and playing with the pellet as fixed effects, with Subject ID as a random effect. Adding time to the model did not improve the fit. In this final model, performance was significantly more accurate in both the positive and the control conditions compared to the negative condition, increased rates of playing with the pellet and self-licking behavior related to better performance, and increased scratching was related to a decrease in performance. Estimates (β) and p values for all of the fixed effects in the final model can be found in Table 3. Overall, the final model reveals that although condition alone significantly predicts accuracy better than the null model, adding behaviors into the model as predictive factors provides a significantly better model for predicting accuracy than condition alone. This suggests to us that it is not just one aspect of the behavior or the manipulations that is driving the results, but that both condition and behavior are helping to explain some of the variance seen in performance. We hypothesize that the additional variance captured by the inclusion of behavior in our model is the result of changes in affective states within the subjects outside of the experimentally elicited condition (i.e., increased scratching in the positive or control condition may have been the result of unforeseen negative stimuli in the subject's environment during a

Figure 4
The Effects of Condition and Time on Number of Trials Completed



Note. Error bars reflect ±1 *SE*.

Figure 5
The Effects of Condition and Time on Accuracy



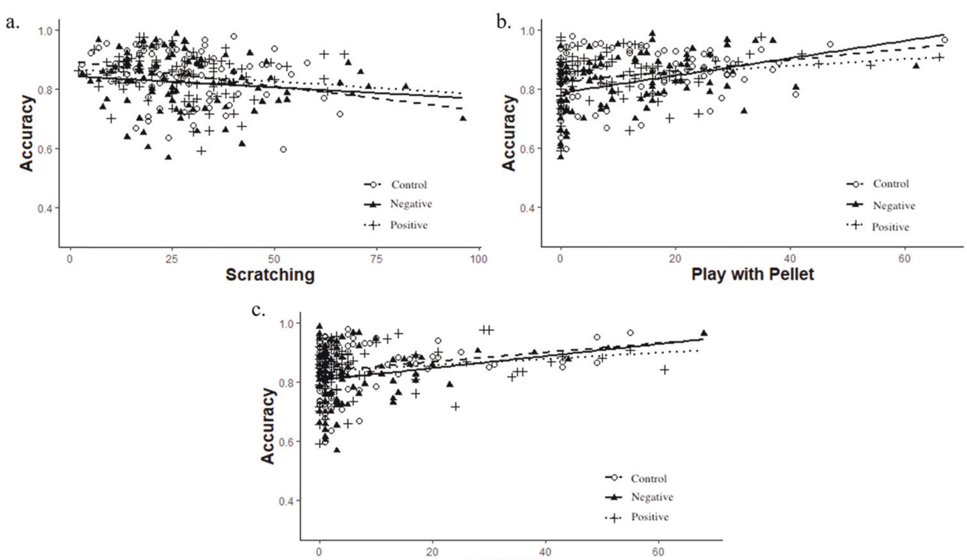
Note. Accuracy measured by the proportion of correct trials to the total number of trials completed. Error bars reflect ±1 *SE*.

particular session), as well as individual differences within the sessions (i.e., subjects with higher levels of negative affect displayed increased scratching and further impairment on the cognitive task).

Discussion

Experiences immediately prior to cognitive testing, intended to alter subjects' affective states, affected both the subsequent behavior and cognitive performance of capuchin monkeys. In line with predictions, a negative experience predicted to be frustrating increased rates of scratching and impaired subsequent performance on a working memory task compared to a positive experience and a control. A second goal of our task was to find behavioral correlates of positive affect. The only behavior that was significantly different in the positive condition as compared to the negative or control conditions was self-licking behavior, which is

Figure 6
Effects of Behavior on Acccuracy



Note. Figure depicts the effects of scratching (a), playing with pellets (b) and self-licking (c) on accuracy (as measured by the proportion of correct trials to the total number of trials completed). Each point represents a single session for each subject.

Table 2
Comparison of the Models Predicting Accuracy on the Delayed Match-to-Sample Task

Models with behavior	*df*	AIC	χ^2	*p*
Null	3	−740.53		
Scratch only	4	−745.19	6.66	.010*
Scratch + Condition	6	−752.36	11.16	.004*
Scratch × Condition	8	−748.40	0.04	.981
Play w/pellet only	4	−751.58	13.05	<.001*
Play w/pellet + Condition	6	−759.57	11.99	.002*
Play w/pellet × Condition	8	−755.94	0.37	.832
Self-Lick only	4	−752.90	14.37	<.001*
Self-Lick + Condition	6	−761.27	12.27	.002*
Self-Lick × Condition	8	−761.01	3.74	.153
Full model	***df***	**AIC**	χ^2	***p***
Null		−1,417.4		
Condition only	2	−1,427.9	14.56	<.001*
Condition + Behaviors	2	−1,465.5	43.53	<.001*
Condition + Behaviors + Time	3	−1,462.1	0.65	.724

* $p < .05$.

challenging to interpret. In addition, contrary to our predictions, the positive condition did not have an augmenting effect on performance on the cognitive task compared to the control. Together, these suggest that our positive experience was not perceived as positive by the monkeys. We discuss each of these points in more detail in the following text.

Our subjects showed increased rates of scratching behavior subsequent to sessions in which the tube was blocked on 40% of the trials (negative condition), compared to following either a positive experience or the control (no tube). Stereotypic behavior also increased in the negative condition compared to the control, but not the positive condition. Increases in these types of behaviors are consistent with our predictions, which were based on the literature on behavioral indicators of negative affective states in primates. Indeed, scratching and stereotypic behaviors are widely used as markers of stress in the primate literature (Baker & Aureli, 1997; Castles et al., 1999; Lutz, 2014; Maestripieri et al., 1992; Norscia & Palagi, 2011; Sclafani et al., 2012; Troisi, 2002). Although it has been argued that parasite load is a greater predictor of scratching than stress in wild primates (Duboscq et al., 2016), our captive monkeys are tested annually for a variety of parasites and did not have any at the time of testing, nor would that explain the variation across conditions. In addition, although we measured other behaviors that have been used previously to indicate negative affect, including self-directed behaviors (self-grooming and self-touching) and yawning (of which we saw no occurrences), none of these were significantly correlated with the negative experience. This suggests that scratching is perhaps the most generalizable and reliable behavioral indicator of negative affect and may function as a manipulation check in future research, supporting the prediction in the current study that the blocked tube was a negative experience for our subjects.

Furthermore, the results from the cognitive aspect of the study point to a clear difference between the negative condition and the other two. As predicted based on previous research on stress in animals and humans (Arnsten & Goldman-Rakic, 1998; Cazakoff et al., 2010; Kuhlmann et al., 2005; Luethi et al., 2009), there was a detrimental effect of a prior negative experience on subsequent accuracy on a DMTS task. As with the behavioral changes, this suggests that our manipulation successfully induced a negative affective state. Although it is possible that lower accuracy on the DMTS could have accounted for the increased scratching, in either case a negative event (the manipulation or the low accuracy) was the cause (and in fact it could also be a combination). This is some of the first evidence that even a mildly frustrating experience can have real consequences for subsequent performance on a cognitive task in primates. This is not only useful in interpreting differences in outcomes across studies that use different stressors, but also important for animal welfare considerations. In addition, our results indicate

Table 3
Fixed Effects Values for the Full Model Predicting Accuracy on the Delayed Match-to-Sample

Fixed effect	Estimate (â)	*SE*	*df*	*t* value	*p*
(Intercept)	8.07×10^{-1}	8.07×10^{-2}	24.6	25.36	<.001*
Control	2.59×10^{-2}	8.07×10^{-3}	769.3	3.24	.001*
Positive	2.04×10^{-2}	8.07×10^{-3}	770.0	3.54	.011*
Scratch	-1.47×10^{-3}	8.07×10^{-4}	782.4	−2.51	.012*
Play w/pellet	3.70×10^{-3}	8.07×10^{-4}	773.8	3.80	<.001*
Self-lick	4.06×10^{-3}	8.07×10^{-3}	717.8	3.45	<.001*

Note. The full model included scratching, self-licking, play with pellet behavior, and condition as fixed effects, with subject ID as a random effect.
* $p < .05$.

that negative experiences prior to testing may influence cognitive performance on subsequent cognitive tasks, which suggests researchers should consider and report experiences prior to data collection that may introduce confounding effects.

Only one recorded behavior, self-licking, correlated with just the positive condition, but we have no explanation for why this might be the case, and we have no evidence of the predicted enhancing effect of positive affect on subsequent cognitive performance. Indeed, if anything, the positive condition seems to have reduced motivation, as subjects completed fewer trials in the first 10 min of the DMTS following the positive condition (but not the second- and third-time blocks), which cannot easily be explained by the presence of food reward (which was the same in the positive and negative conditions) or the presence of frustration (as they behaved the same in the control and the negative conditions). Perhaps they were either calmer, and therefore moved more slowly, or were less motivated to seek out additional reward after a positive experience. Taken together, these inconclusive findings combined with the lack of other behavioral or cognitive effects suggest that we may not have induced positive affect. There are several possible explanations for this. For instance, we cannot control the experiences prior to testing for our socially housed subjects or even the behavior of conspecifics nearby *during* testing. However, these experiences would be highly variable, and the consistency of our negative impact suggests that these either were not strong effects (or at least not strong enough to override the negative experience) or had no effect at all. Alternatively, it is possible that the monkeys did not find the task rewarding, despite receiving the high value food reward. Indeed, some subjects had poor technique; they never learned to pull the slide straight up, instead trying to pull it toward themselves, which made the task substantially harder. Again, however, even those who had difficulty responded more negatively when rewards were unobtainable. It is also possible that we did induce positive affect, but that there were no behavioral correlates and that our cognitive test was not sufficiently sensitive to detect it. Some research with humans indicates that although there may be an overall improving effect of experiencing positive affect on working memory, this effect is stronger for more complicated tasks (an operation-span task compared to a word-span task; Yang et al., 2013).

We hypothesize, however, that the explanation is either that the positive condition was not sufficiently positive, or that any positive affect was overridden by subjects' frustration when the apparatus was removed, prior to the computerized DMTS. Indeed, the most common finding from our behavioral analysis was that behaviors either differed only for the negative condition or differed for the negative and positive conditions relative to the control. One possible explanation is that in the experimental conditions, the monkeys received six pieces of cereal, but received none in the control. Attributing the significant behavioral differences to satiation effects seems unwarranted, however, as our monkeys can easily eat dozens of pieces of cereal and presumably were not greatly impacted by six small pieces. The other difference between the experimental and control conditions was the presence of the apparatus or, more specifically, its removal prior to the DMTS test. Anecdotally, the monkeys did not like that; some subjects would occasionally vocalize and threaten the experimenter as the tube was being unclipped from their test box, and these monkeys would often grab on to the apparatus and try to prevent its removal, sometimes to the point of nearly (or actually) breaking it. In such cases, the experimenter released their hold until the monkey lost interest and then proceeded with the removal. Although clearly this did not totally override the experience, as subjects still responded differently in the negative condition, it could have been sufficient to wipe out any positive affect that we generated. Whether or not we actually generated positive affect, this reiterates the importance of using similar paradigms for both positive and negative manipulations in future research; if we had used a nonmanipulative paradigm for the positive condition, we could have mistaken these behavioral differences that changed in both conditions as behavioral indicators of negative affect rather than recognizing them as a confounding variable.

It is important to note that our final, best-fit model for predicting accuracy on the task included both condition and behavior as predictive factors. In other words, including behavior in the model significantly improved predictions of performance *beyond* what condition alone predicted. Although we hypothesize that this could be in part because the behavioral results are picking up some of the variance caused by differences in levels of affect and affective changes outside of the experimental conditions, we cannot determine that with the current data. Indeed, one target of future study should be whether further differences in the rates of certain behaviors are indicative of changes in the varying intensity of affective states.

Finally, although finding that cognitive performance is impaired in a nonhuman species following even a mildly negative experience is an exciting result, we cannot speak to mechanism. Future research will hopefully clarify the relationship and causal connection between negative experiences and cognitive performance, as well as the role of behavior, and expand this research to more species. Perhaps more importantly, we hope future research will pinpoint noninvasive ways to measure positive affect in numerous animal species, which will improve not only our understanding of the impact of affect on cognition and behavior and our ability to study positive affect generally, but also provide a method for measuring and improving animals' welfare.

References

Allritz, M., Call, J., & Borkenau, P. (2016). How chimpanzees (Pan troglodytes) perform in a modified emotional Stroop task. *Animal Cognition*, *19*(3), 435–449. https://doi.org/10.1007/s10071-015-0944-3

Amici, F., Barney, B., Johnson, V. E., Call, J., & Aureli, F. (2012). A modular mind? A test using individual data from seven primate species. *PLoS ONE*, *7*(12), Article e51918. https://doi.org/10.1371/journal.pone.0051918

Arnsten, A. F. T., & Goldman-Rakic, P. S. (1998). Noise stress impairs prefrontal cortical cognitive function in monkeys: Evidence for a hyperdopaminergic mechanism. *Archives of General Psychiatry*, *55*(4), 362–368. https://doi.org/10.1001/archpsyc.55.4.362

Ashby, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review*, *106*(3), 529–550. https://doi.org/10.1037/0033-295x.106.3.529

Baker, K. C., & Aureli, F. (1997). Behavioural indicators of anxiety: An empirical test in chimpanzees. *Behaviour*, *134*(13–14), 1031–1050. https://doi.org/10.1163/156853997X00386

Barrett, L. F., & Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Advances in Experimental Social Psychology*, *41*, 167–218. https://doi.org/10.1016/S0065-2601(08)00404-8

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beran, M. J., Evans, T. A., Klein, E. D., & Einstein, G. O. (2012). Rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella) remember future responses in a computerized task. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*(3), 233–243. https://doi.org/10.1037/a0027796

Beran, M. J., Klein, E. D., Evans, T. A., Chan, B., Flemming, T. M., Harris, E. H., Washburn, D. A., & Rumbaugh, D. M. (2008). Discrimination reversal learning in capuchin monkeys (Cebus apella). *The Psychological Record*, *58*(1), 3–14. https://doi.org/10.1007/BF03395599

Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella). *Cognition*, *120*(1), 90–105. https://doi.org/10.1016/j.cognition.2011.02.016

Bethell, E. J., Holmes, A., Maclarnon, A., & Semple, S. (2012). Evidence that emotion mediates social attention in rhesus macaques. *PLoS ONE*, *7*(8), Article e44387. https://doi.org/10.1371/journal.pone.0044387

Boissy, A., Manteuffel, G., Jensen, M. B., Moe, R. O., Spruijt, B., Keeling, L. J., Winckler, C., Forkman, B., Dimitrov, I., Langbein, J., Bakken, M., Veissier, I., & Aubert, A. (2007). Assessment of positive emotions in animals to improve their welfare. *Physiology and Behavior*, *92*(3), 375–397. https://doi.org/10.1016/j.physbeh.2007.02.003

Brodeur, M. B., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. *PLoS ONE*, *9*(9), Article e106953. https://doi.org/10.1371/journal.pone.0106953

Castles, D. L., Whiten, A., & Aureli, F. (1999). Social anxiety, relationships and self-directed behaviour among wild female olive baboons. *Animal Behaviour*, *58*(6), 1207–1215. https://doi.org/10.1006/anbe.1999.1250

Cazakoff, B. N., Johnson, K. J., & Howland, J. G. (2010). Converging effects of acute stress on spatial and recognition memory in rodents: A review of recent behavioural and pharmacological findings. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*(5), 733–741. https://doi.org/10.1016/j.pnpbp.2010.04.002

Celli, M. L., Tomonaga, M., Udono, T., Teramoto, M., & Nagano, K. (2003). Tool use task as environmental enrichment for captive chimpanzees. *Applied Animal Behaviour Science*, *81*(2), 171–182. https://doi.org/10.1016/S0168-1591(02)00257-5

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298*(5596), 1191–1194. https://doi.org/10.1126/science.1076358

Duboscq, J., Romano, V., Sueur, C., & MacIntosh, A. J. J. (2016). Scratch that itch: Revisiting links between self-directed behaviour and parasitological, social and environmental factors in a free-ranging primate. *Royal Society Open Science*, *3*(11), Article 160571. https://doi.org/10.1098/rsos.160571

Duncan, S., & Barrett, L. F. (2007). Affect is a form of cognition: A neurobiological analysis. *Cognition and Emotion*, *21*(6), 1184–1211. https://doi.org/10.1080/02699930701437931

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. https://doi.org/10.3758/brm.40.2.590

Fragaszy, D. M., Visalberghi, E., & Fedigan, L. M. (2004). *The complete capuchin: The biology of the genus Cebus*. Cambridge University Press.

Frost, R. O., & Green, M. L. (1982). Velten mood induction procedure effects duration and postexperimental removal. *Personality and Social Psychology Bulletin*, *8*(2), 341–347. https://doi.org/10.1177/0146167282082024

Garner, J. P. (2005). Stereotypies and other abnormal repetitive behaviors: Potential impact on validity, reliability, and replicability of scientific outcomes. *ILAR Journal*, *46*(2), 106–117. https://doi.org/10.1093/ilar.46.2.106

Garner, J. P., Meehan, C. L., & Mench, J. A. (2003). Stereotypies in caged parrots, schizophrenia and autism: Evidence for a common mechanism. *Behavioural Brain Research*, *145*(1–2), 125–134. https://doi.org/10.1016/S0166-4328(03)00115-3

Gomez, P., Zimmermann, P. G., Guttormsen Schär, S., & Danuser, B. (2009). Valence lasts longer than arousal: Persistence of induced moods as assessed by psychophysiological measures. *Journal of Psychophysiology*, *23*(1), 7–17. https://doi.org/10.1027/0269-8803.23.1.7

Hagen, K., & Broom, D. M. (2004). Emotional reactions to learning in cattle. *Applied Animal Behaviour Science*, *85*(3), 203–213. https://doi.org/10.1016/j.applanim.2003.11.007

Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *The Journal of Neuroscience*, *25*(11), 2977–2982. https://doi.org/10.1523/JNEUROSCI.5139-04.2005

Lacreuse, A., Schatz, K., Strazzullo, S., King, H. M., & Ready, R. (2013). Attentional biases and memory for emotional stimuli in men and male rhesus monkeys. *Animal Cognition*, *16*(6), 861–871. https://doi.org/10.1007/s10071-013-0618-y

Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes*, *117*, 52–58. https://doi.org/10.1016/j.beproc.2014.11.019

Luethi, M., Meier, B., & Sandi, C. (2009). Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. *Frontiers in Behavioral Neuroscience*, *2*, Article 5. https://doi.org/10.3389/neuro.08.005.2008

Lutz, C. K. (2014). Stereotypic behavior in nonhuman primates as a model for the human condition. *ILAR Journal*, *55*(2), 284–296. https://doi.org/10.1093/ilar/ilu016

Maestripieri, D., Schino, G., Aureli, F., & Troisi, A. (1992). A modest proposal: Displacement activities as an indicator of emotions in primates. *Animal Behaviour*, *44*(5), 967–979. https://doi.org/10.1016/S0003-3472(05)80592-5

Marzouki, Y., Gullstrand, J., Goujon, A., & Fagot, J. (2014). Baboons' response speed is biased by their moods. *PLoS ONE*, *9*(7), Article, e102562. https://doi.org/10.1371/journal.pone.0102562

McGowan, R. T. S., Rehn, T., Norling, Y., & Keeling, L. J. (2014). Positive affect and learning: Exploring the "Eureka Effect" in dogs. *Animal Cognition*, *17*(3), 577–587. https://doi.org/10.1007/s10071-013-0688-x

McGowan, R. T. S., Robbins, C. T., Alldredge, J. R., & Newberry, R. C. (2010). Contrafreeloading in grizzly bears: Implications for captive foraging enrichment. *Zoo Biology*, *29*(4), 484–502. https://doi.org/10.1002/zoo.20282

Meehan, C. L., & Mench, J. A. (2007). The challenge of challenge: Can problem solving opportunities enhance animal welfare? *Applied Animal Behaviour Science*, *102*(3–4), 246–261. https://doi.org/10.1016/j.applanim.2006.05.031

Norscia, I., & Palagi, E. (2011). When play is a family business: Adult play, hierarchy, and possible stress reduction in common marmosets. *Primates*, *52*(2), 101–104. https://doi.org/10.1007/s10329-010-0228-0

Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience*, *9*, Article 58. https://doi.org/10.3389/fnhum.2015.00058

Panksepp, J. (2004). *Affective neuroscience: The foundations of human and animal emotions*. Oxford University Press.

Paul, E. S., Harding, E. J., & Mendl, M. (2005). Measuring emotional processes in animals: The utility of a cognitive approach. *Neuroscience and Biobehavioral Reviews*, *29*(3), 469–491. https://doi.org/10.1016/j.neubiorev.2005.01.002

Pomerantz, O., Terkel, J., Suomi, S. J., & Paukner, A. (2012). Stereotypic head twirls, but not pacing, are related to a 'pessimistic'-like judgment bias among captive tufted capuchins (Cebus apella). *Animal Cognition*, *15*(4), 689–698. https://doi.org/10.1007/s10071-012-0497-7

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rilling, J. K., & Insel, T. R. (1999). The primate neocortex in comparative perspective using magnetic resonance imaging. *Journal of Human Evolution*, *37*(2), 191–223. https://doi.org/10.1006/jhev.1999.0313

Schoofs, D., Preuss, D., & Wolf, O. T. (2008). Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology*, *33*(5), 643–653. https://doi.org/10.1016/j.psyneuen.2008.02.004

Sclafani, V., Norscia, I., Antonacci, D., & Palagi, E. (2012). Scratching around mating: Factors affecting anxiety in wild Lemur catta. *Primates*, *53*(3), 247–254. https://doi.org/10.1007/s10329-012-0294-6

Seed, A., & Tomasello, M. (2010). Primate cognition. *Topics in Cognitive Science*, *2*(3), 407–419. https://doi.org/10.1111/j.1756-8765.2010.01099.x

Shibasaki, M., & Kawai, N. (2009). Rapid detection of snakes by Japanese monkeys (Macaca fuscata): An evolutionarily predisposed visual system. *Journal of Comparative Psychology*, *123*(2), 131–135. https://doi.org/10.1037/a0015095

Tavares, M. C. H., & Tomaz, C. (2002). Working memory in capuchin monkeys (Cebus apella). *Behavioural Brain Research*, *131*(1–2), 131–137. https://doi.org/10.1016/S0166-4328(01)00368-0

Troisi, A. (2002). Displacement activities as a behavioral measure of stress in nonhuman primates and human subjects. *Stress*, *5*(1), 47–54. https://doi.org/10.1080/102538902900012378

Truppa, V., De Simone, D. A., Piano Mortari, E., & De Lillo, C. (2014). Effects of brief time delays on matching-to-sample abilities in capuchin monkeys (Sapajus spp.). *Behavioural Brain Research*, *271*, 240–248. https://doi.org/10.1016/j.bbr.2014.05.023

Yang, H., Yang, S., & Isen, A. M. (2013). Positive affect improves working memory: Implications for controlled cognitive processing. *Cognition and Emotion*, *27*(3), 474–482. https://doi.org/10.1080/02699931.2012.713325

Appendix

List of Recorded Behaviors

Table A1

All Behaviors Recorded for Behavioral Analysis

Behavior	Note
Scratch	
Play with pellet	
Threaten	
Self-lick	
Self-touch	
Lick-cage	Combined into “Cage-Directed Behavior”
Pick at cage	Combined into “Cage-Directed Behavior”
Wipe-cage	Combined into “Cage-Directed Behavior”
Head-twirl	Combined in “Stereotypic Behavior”
Rub-hands	Combined in “Stereotypic Behavior”
Pace	Combined in “Stereotypic Behavior”
Shake	
Urine wash	
Auto-groom	
Reach	
Lip-smack	Observed in less than ten subjects
Push faceplate	Observed in less than ten subjects
Manipulate	Observed in less than ten subjects
Play with water	Observed in less than ten subjects

Note. Every occurrence of behavior was recorded. Behavior was counted as a new occurrence instead of a continuation of the last if there was a 3-s period between the end of the first occurrence and the start of the second.

Received August 4, 2020
Revision received March 4, 2021
Accepted March 9, 2021 ■