

# High emotional reactivity toward an experimenter affects participation, but not performance, in cognitive tests with common marmosets (*Callithrix jacchus*)

Michèle N. Schubiger · Florian L. Wüstholtz ·  
André Wunder · Judith M. Burkart

Received: 7 October 2014 / Revised: 20 December 2014 / Accepted: 5 January 2015 / Published online: 31 January 2015  
© Springer-Verlag Berlin Heidelberg 2015

**Abstract** When testing primates with cognitive tasks, it is usually not considered that subjects differ markedly in terms of emotional reactivity toward the experimenter, which potentially affects a subject's cognitive performance. We addressed this issue in common marmosets (*Callithrix jacchus*), a monkey species in which males tend to show stronger emotional reactivity in testing situations, whereas females have been reported to outperform males in cognitive tasks. In a two-phase experiment, we first quantified the emotional reactivity of 14 subjects toward four different experimenters performing a standardized behavioral action sequence and then assessed whether and how it affected the subjects' participation and performance in a subsequent object permanence task. A test session was terminated if a subject refused to make a choice in four consecutive trials. Highly emotionally aroused individuals, particularly males, were less likely to participate in the cognitive task and completed fewer trials. However, whenever they did participate and were attentive to the

task, their performance was not affected. Our results suggest that differences in emotional reactivity toward an experimenter have no major impact on cognitive performance if strict criteria are applied on when to abandon a test session and if performance is corrected for attention to the test procedure. Furthermore, they suggest that the reported sex differences in cognitive performance in marmosets may be owing to motivational and attentional factors, rather than a difference in cognitive ability per se.

**Keywords** Emotional reactivity · Common marmosets · Experimenter effects · Sex differences in cognitive performance

## Introduction

Whenever researchers conduct experiments with human or non-human primates, there is a risk that they will unintentionally influence the study's outcome in one or several ways. Consequently, the study subjects might change their behavior in response to the experimenter rather than the experimental conditions. In order to avoid experimenter effects and to optimize comparability of results, different experimenters are usually trained to use identical standardized procedures. But can we ever be certain that different experimenters conducting the same test with the same subjects reliably obtain the same results in terms of the subjects' cognitive performance scores? One possibility is that different experimenters may elicit different emotional reactions in the subjects. In particular, some subjects may feel more at ease or more nervous with a specific experimenter than others. Such individual differences in emotional reactivity may influence a subject's motivation to participate in a cognitive test with this experimenter,

---

**Electronic supplementary material** The online version of this article (doi:10.1007/s10071-015-0837-5) contains supplementary material, which is available to authorized users.

---

M. N. Schubiger (✉) · J. M. Burkart  
Anthropological Institute and Museum, University of Zurich,  
Winterthurerstrasse 190, 8057 Zurich, Switzerland  
e-mail: mnschubiger@gmail.com; Michele.Schubiger@uzh.ch

F. L. Wüstholtz  
Department of Philosophy, University of Fribourg,  
Av. Europe 20, 1700 Fribourg, Switzerland

A. Wunder  
Department of Theoretical Philosophy, University of Zurich,  
Zürichbergstrasse 43, 8044 Zurich, Switzerland

and, importantly, this lower motivation in turn may affect its cognitive performance.

Possible effects of experimenters on subjects' performance scores in cognitive studies have been extensively studied in humans, and rodents, but only rarely been addressed systematically in non-human primates. The numerous reports of experimenter effects on the cognitive performance of human subjects include experimenters forming personal expectations on how different subjects might perform (e.g., Rosenthal 1963) or differences in the way experimenters interact with their subjects such as being friendly or neutral (e.g., Siegwarth et al. 2012). In studies with rats, biases such as an experimenter's expectation regarding a subject's cognitive abilities or the degree of familiarity between the experimenter and the subject have been shown to potentially affect a subject's performance (see Schellinck et al. 2010). Systematic investigations of experimenter effects in non-human primates are rare and include two studies on capuchin monkeys, an observational field study on the putative effects of the presence of a human observer on the subjects' movement and activity patterns (Crofoot et al. 2010) and an experimental study that suggested experimenters may bias the results of cognitive tests by selectively choosing subjects with homogenous personality traits (Morton et al. 2013).

Crofoot et al. (2010) analyzed the movement and activity patterns (recorded via telemetric collars) of seven habituated capuchin monkeys (*Sapajus apella*) in the presence and absence of a human researcher. Overall, the capuchins did not change their movement and activity patterns in the presence of a human researcher. More subtle behavioral differences, however, could not be measured, such as the relative frequency of particular activities (e.g., vigilance, distress or resting behavior) and whether individuals differ systematically in such responses to the presence of a human researcher.

Morton et al. (2013) analyzed whether the outcome of cognitive tests in captivity is affected by the so-called personality selection bias. In captive experiments, researchers can control for most confounding variables by careful study design. However, researchers may tend to selectively choose their study subjects based on particular personality traits, such as openness. This selection bias may ultimately result in a performance bias if such a personality trait is linked to cognitive performance. Morton and colleagues rated capuchin monkeys (*S. apella*) on five personality dimensions and analyzed their participation (how many sessions they completed) and performance (how fast they learned to choose correctly) in two training tasks. In the first task, food was positioned in front of one of the two test compartments that the monkeys could freely access, and a correct response was to sit in the compartment that had the food in front of it. In the second task, a

small opaque cup was placed in front of one compartment and a larger opaque cup in front of the other one, and a correct response was to sit in the compartment that had the larger cup in front of it. The authors found that individuals scoring higher on *openness* were more likely to participate and performed better in the first task, and *assertiveness* affected the monkeys' performance in both tasks, with less assertive subjects performing better than more assertive ones. Arguably, only the second task measured a truly cognitive component, and it is in this task that an individual's openness score no longer had an effect on its performance score. The negative effect of assertiveness may be best understood as individuals having motivational priorities others than food. A direct link between personality traits and cognitive ability per se could thus not be demonstrated unambiguously.

A study on rhesus macaques (*Macaca mulatta*, Toxopeus et al. 2005), however, showed that high levels on one dimension of trait anxiety, as assessed behaviorally in the group context in response to a loud noise, negatively impacted performance in a learning test. This result may suggest a link between one of the three dimensions of trait anxiety and cognitive performance, but, as the authors stress, highly anxious subjects could simply be less attentive to the test than less anxious subjects, rather than having genuinely lower cognitive abilities. This argument is particularly convincing since anxiety was also linked with low status: Subordinate individuals may be less attentive to the task simply because they have to constantly monitor dominant conspecifics. In order to disentangle such confounding effects from genuine cognitive ability, it is necessary to control cognitive performance with regard to whether the subjects were attentive to the test procedure in the first place or not.

Importantly, excluding subjects from participation in cognitive tests is problematic, regardless of whether exclusions target individuals of a specific personality type, to the extent that they differ systematically in genuine cognitive performance. But even in the absence of such sampling biases, the risk of indirect and subtle experimenter influences on the subjects' performance remains. According to the Yerkes–Dodson law (Yerkes and Dodson 1908), an individual performs best in a cognitive task if its emotional arousal is at an optimal level, which tends to be a medium degree of arousal. This optimal arousal level may vary both between and within individuals. While personality and trait anxiety may to some extent explain the origin of arousal level differences between individuals, an individual's emotional arousal levels also may vary from one cognitive task to the other, either as a consequence of task difficulty or a variety of other causes. Therefore, it is essential that the experimenter is aware of individual differences in the subjects' emotional reactivity and their

possible consequences on the subjects' motivation, attention and cognitive performance. Ideally, a subject's behavioral signs of emotional arousal are evaluated prior to and during cognitive testing. The aim of our study was to systematically assess these issues in common marmoset monkeys (*Callithrix jacchus*).

Marmosets belong to the primates smallest in body size (Ford et al. 2009), which arguably makes them particularly likely to show high emotional reactivity toward human experimenters owing to the body size difference between experimenter and subject. Furthermore, reported and observed sex differences make them a particularly interesting study species to investigate the effect of emotional reactivity toward the experimenter on their participation and performance in cognitive tasks. In the wild, male marmosets are more vigilant than females (Koenig 1998), and in captivity, they seem to be more easily emotionally aroused during cognitive testing (personal observation). Emotionally aroused males often appear less willing to participate in a full test session or less attentive to the experimental stimuli when they do so. Intriguingly, male marmosets have also been reported to perform more poorly in cognitive tasks than females (Brown et al. 2010; Yamamoto et al. 2004). They thus represent an ideal test case to identify whether their poor performance simply reflects emotional reactivity, a lack of motivation and attention or a true sex difference in cognitive ability. Yamamoto et al. (2004) tested female and male marmosets with a cognitive task that required the subjects to open an opaque plastic container by piercing its lid in order to obtain a raisin that they could see through a small hole in the lid. Male marmosets needed more time to approach and solve this task and used fewer strategies to obtain the reward. Brown et al. (2010) tested female and male marmosets' capacity to perceive biological motion by presenting them with point-light patterns of a walking hen on a computer screen. They found that female subjects paid more attention to the biological motion stimuli than males and spent more time inspecting them. The results of both above-mentioned studies were interpreted as a sex differences in marmosets' cognitive abilities. However, an alternative explanation is that the poor performance of male marmosets found in these studies was an artifact of a lower motivation (e.g., to work for food) or lower attention resulting from higher emotional arousal in the test situation.

The aim of our study was to systematically assess individual differences in the subjects' spontaneous behavioral response to an experimental situation and their potential effect on their participation and performance in a cognitive task. In a two-phase experiment conducted in a single session, we quantified behavioral and bodily indicators of elevated emotional reactivity when marmosets were confronted with different experimenters and investigated

whether variation in emotional reactivity toward an experimenter influenced the subjects' motivation to participate and their performance in a commonly used cognitive task.

In the first phase, an experimenter performed a standardized sequence of behavioral actions in front of the individual subject. This sequence consisted of the experimenter entering the room, performing a succession of six standardized actions, leaving and reentering the room and finally offering a highly desirable reward to the subject. In the second phase, the same experimenter conducted an object permanence task with the subject, composed of a visible and an invisible displacement condition of six trials each. In order to maximize variation in emotional reactivity, we used different experimenters that were either familiar women (whom we expected to elicit low levels of emotional arousal) or unfamiliar men (whom we expected to elicit high levels of emotional arousal).

We addressed three main questions: (1) Do individual subjects react differently to different experimenters in terms of emotional arousal? (2) Are subjects with a stronger emotional response to an experimenter less willing to participate in a cognitive test? (3) Do subjects with a stronger emotional response toward an experimenter perform worse in a cognitive task given they participate?

We expected to find substantial variation in the subjects' emotional reactivity and that male marmosets would react more strongly, particularly in response to unfamiliar experimenters. Furthermore, we anticipated that high emotional reactivity would lead to low participation, whereas we had no prediction on whether high emotional reactivity would be linked to cognitive performance given we controlled for attention to the test stimuli and their manipulation. If males show higher emotional reactivity and lower participation as well as lower performance, even after controlling for attention, this would suggest genuine sex differences in cognitive ability. If, however, they showed higher reactivity, lower participation, but no difference in cognitive performance after controlling for attention, this would suggest that sex differences in cognitive performance are driven by motivational and attentional factors, rather than cognitive ability per se.

## Materials and methods

### Subjects

Study subjects were 14 adult captive common marmosets (*C. jacchus*) of equal sex distribution, with a mean age of 7.11 years, living in family groups at the Primate Station of the Anthropological Institute of the University of Zurich. All marmosets were captive-born and mother-reared and participated in this study on four mornings after being fed

their regular breakfast meal, a vitamin and calcium-enriched porridge. For a description of the study sample, see Table 1.

### Experimenters

To maximize variation in the marmosets' emotional reactivity, we used four different human experimenters: two familiar women whom we expected to elicit low emotional arousal and two unfamiliar men whom we expected to elicit high emotional arousal based on our experience with the study species (similar effects have recently been reported for rodents by Sorge et al. 2014). The marmosets had never seen the unfamiliar experimenters before this study but regularly participated in other studies with the familiar experimenters for at least a year. Note that the aim of this study was not to identify why marmosets react more strongly to some experimenters than others, i.e., to disentangle the effect of the experimenters' familiarity or gender on marmosets' emotional reactivity, which would have required counterbalancing the two factors. Rather, we were interested in the consequences of high emotional reactivity, regardless of its origin, for the marmosets' participation and performance in a subsequent cognitive task. Consequently, we chose experimenters whom we anticipated to produce the highest variation in the subjects' emotional reactivity. All four experimenters wore indiscernible black clothes and, most importantly, featureless black shoes as marmosets, an arboreal New World monkey species, tend to react strongly to obtrusive features at ground level (unpublished data).

### Experimental setup

All marmosets were tested individually in the same familiar experimental room and cage in which they had

previously been tested in several socio-cognitive experiments (e.g., Burkart et al. 2007, 2009). The experimental cage contained the actual test compartment (41 × 33 × 53 cm) and a directly connected second compartment of the same size. The two compartments were divided by a grid wall containing a rectangular opening (15 × 22.5 cm) in its bottom half through which the subject could move away from the experimenter and testing table and closer to its group members. We thereby ensured that very highly aroused subjects did not feel restricted in space. Each subject was tested individually in the same test compartment while its group members were waiting in an adjacent cage (100 × 78 × 122 cm). This waiting cage was connected to the experimental cage through a short gateway. Its lateral wall facing the experimental enclosure was covered with an opaque gray plastic board and its front with an opaque cream-colored fabric blind. This setup prevented the waiting subjects from seeing the experimenter until it was their turn to enter the test compartment while still allowing the current subject to hear and smell, but not see, its group members. Having such minimal "contact" with group members is essential for marmosets' welfare as complete isolation from conspecifics imposes distress on this highly social species. Both cages were elevated 1.15 m from the ground as marmosets are arboreal and thus appear more comfortable in higher positions.

### Apparatus

The basic experimental apparatus consisted of a testing table, a wooden board (40 × 40 cm) with two lateral plastic rails that was mounted on an adjustable tripod table so that the board's front was flush with the front of the experimental compartment. The wooden board of the testing table served as a base for the experimental equipment during the experimenter's behavioral action sequence (phase 1) and for the sliding platform and the cups during the cognitive task (phase 2). The equipment for the experimenter's behavioral action sequence consisted of two transparent glass containers (6 × 3 cm) with white plastic lids, one of which was half-filled with fine-grained pet sand, as well as of a digital timer and a dead cricket (*Acheta domesticus*; about 12 mm in size). The cognitive test apparatus itself consisted of a rectangular sliding platform (33.7 × 25 cm) made of cardboard that was covered with a washable wood-patterned plastic surface and three cylindrical opaque black plastic cups (3.8 × 2.6 cm). The three aligned cups were placed at a distance of 3 cm from the platform's front edge, and the distance between the middle and each outer cup was 10 cm. All three cups were initially openly placed on the sliding board and kept in place with double-sided adhesive pads attached to the side that was lying on the board.

**Table 1** Description of the study sample

Subject	Sex	Age
Jugo	Male	5.8
Juri	Male	11.0
Kaliper	Male	10.8
Kapi	Male	9.8
Tabor	Male	3.7
Tale	Female	4.3
Tessy	Female	11.5
Thilo	Male	4.3
Venezia	Female	6.5
Venus	Female	7.7
Verona	Female	6.5
Vesta	Female	7.7
Vito	Male	6.0
Vreni	Female	11.0

Listed are the 14 subjects' name, sex (seven females and seven males) and age in years

## Experimental procedure

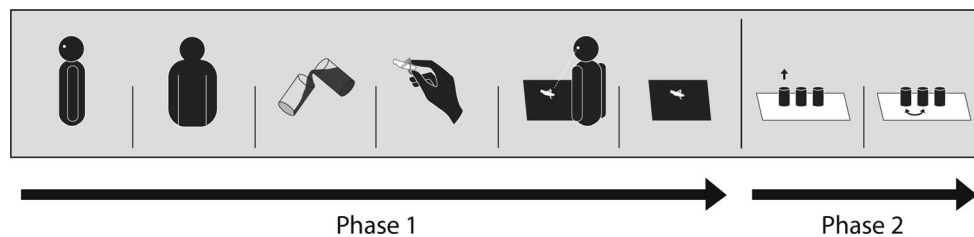
The experimental procedure consisted of two consecutive phases conducted in a single session with a maximal total duration of approximately 15 min that mainly depended on whether and how long an individual subject participated in the cognitive task. We used a within-subject design, where each of the four experimenters performed the two-phase experiment with all subjects. In phase 1, the experimenter (hereafter E) performed a series of six standardized behavioral actions in front of the subject that each lasted 30 s. In phase 2, the E conducted a short object permanence task with the same subject. This resulted in 56 experimenter–subject dyads. Both phases, the standardized behavioral action sequence and the object permanence task, were identical for all subject–experimenter pairs. Experimenter order was counterbalanced across subjects to control for order effects. The two phases of the experiment are shown in Fig. 1.

### *Phase 1: the standardized behavioral action sequence of the experimenter*

The E entered the room, called the subject's name, walked toward the experimental cage on a marked line on the floor while looking down to this line and stopped 50 cm in front of the testing compartment, leaving the testing table between the E and the subject. The E then performed a standardized sequence of six behavioral actions, each lasting 30 s. A digital timer beeped to indicate when a 30-s interval had ended. For each action of the behavioral sequence, the E retrieved the equipment from a storage table at the back wall of the room and replaced it onto this storage table before starting the next action.

### *The experimenters' six sequential actions*

1. *Standing still* The E stood in front of the subject's test compartment. The E's body, head and eyes were facing 90° away from the subject, while the arms were in a neutrally stretched position on the sides of the E's body. After 30 s, the E turned toward the subject and began with the second action.
2. *Establishing eye contact* The E was now facing the subject and tried to establish eye contact for 30 s by following the subject with head and eye gaze.
3. *Manipulation of an unfamiliar object* The E took the two small transparent containers (one was filled with sand and the other was empty) and placed them on the testing table. The E started the timer and removed the containers' lids, subsequently lifted up the containers, one in each hand, and visibly poured the sand from one container into the other and vice versa as soon as the first container was empty. The E repeated pouring the sand between the two containers, until 30 s had elapsed, closed the lids again and placed the containers back onto the storage table.
4. *Holding food* After taking a cricket from the storage table, the E again stood in front of the testing table, as in the previous sequence, started the timer and held up the cricket visibly about 25 cm in front of E's chest without offering it to the subject. While doing so, the E was facing the digital timer on the apparatus for 30 s.
5. *Placing food out of reach/E present* The E laid down the previously held cricket onto the wooden platform at a distance of 17 cm (which is just out of the subject's reach) from the front of the test compartment and looked at the cricket for 30 s while standing still.
6. *Placing food out of reach/E absent* The E left the room while the cricket remained placed on the wooden platform and out of the subject's reach. After the 30-s interval had elapsed, E reentered the room, picked up



**Fig. 1** The two consecutive phases of the experiment. In phase 1, the experimenter (E) performed a sequence of six behavioral actions in front of the subject: (1) standing perpendicularly to it with head and eyes oriented away, (2) looking at the subject while oriented toward it and establishing eye contact, (3) manipulating an object (pouring sand between two transparent containers), (4) holding up a cricket (out of reach), (5) placing the cricket onto the test table/board (out of reach) and (6) leaving the test room (cricket still out of reach), reentering it

and giving the cricket to the subject. In phase 2, the E conducted a cognitive task with the subject that consisted of two short object permanence tests of six trials each. In the first test, E placed the cricket under one of the three cups and closed all cups. In the second test, E again placed the cricket under one cup but then exchanged its location with an adjacent one. The subject made its choice by touching or lifting one of the cups

the out-of-reach cricket from the wooden board/test table and offered it to the subject.

### *Phase 2: object permanence task*

Immediately after completing the behavioral action sequence, the E conducted a short object permanence task with the subject, consisting of two tests of six trials each, (1) a visible displacement test and (2) an invisible displacement test. The procedure corresponded principally to the one used by Mendes and Huber (2004) who found considerable individual variation in marmosets' performance. In the visible displacement test, the E placed the reward twice under each of the three cups in a counter-balanced manner. In the invisible displacement test and in contrast to Mendes and Huber, we did not use a transport cup and ensured that, at the time of the subject's choice, the reward appeared equally often in each location, twice on the left (A), twice in the middle (B) and twice on the right (C). To enhance comparability, all subjects first received six trials of the simpler test, the visible displacement. The invisible displacement test also consisted of six trials and followed immediately afterwards.

**Task procedure** The E started the testing session by first placing the sliding platform with the three open black cups onto the test table. The E then called the subject's name, showed it a cricket, placed it in front of one of the open cups and closed the cups from left to right.

**Test 1: visible displacement** In the visible displacement test, directly after baiting, the E slid the platform toward the test compartment so that the subject was able to make its choice by touching or lifting one of the cups.

**Test 2: invisible displacement** In the invisible displacement condition, the E baited one cup in the same way as in the visible displacement condition but performed a transposition immediately after closing all cups. The transposition consisted in the E exchanging the location of two cups, the baited and an adjacent empty one by simultaneously moving them on the board from one location to the other using the index and middle fingers of both hands. This resulted in the E's hands crossing, whereby the hand moving the baited container was always in the front. As in the visible displacement condition, E then slid the platform toward the test compartment and allowed the subject to make its choice.

For both the visible and the invisible displacement tests, a choice was defined as the first cup the subject either lifted or touched. The procedure following the subject's choice was identical for both tests. If the subject correctly chose the baited cup, it obtained the cricket reward and the next trial started immediately. If the subject chose the wrong

cup, the E drew back the platform, showed the subject the wrongly chosen container was empty and then opened the other two cups from left to right. The E then retrieved the cricket, placed it on the storage table, and continued with the next trial after a 15-s time delay. If the subject refused to make a choice in a given trial within 30 s, the platform was also retrieved but no such time out imposed. The next trial was directly started instead, and the no-choice trial was repeated once all six trials of the test had been conducted, e.g., after the sixth trial of the visible displacement test.

Crucially, clear stop criteria were applied on when to terminate a test session. Testing was discontinued if a subject refused to choose any cup in four consecutive trials or if a subject did no longer make a choice in the last three trials of the invisible displacement test. This procedure ensured that subjects were allowed to leave the test situation if they were highly emotionally aroused, refused to make any choices or stopped choosing during the test session and thus appeared unmotivated. Therefore, while all subjects were presented with at least four of the six trials in the first test (visible displacement), not all subjects entered the second test (invisible displacement).

### *Data recording and analysis*

Both phases of the experiment, the behavioral action sequence of the E and the cognitive task, were video-recorded continuously with a digital camera. The subjects' spontaneous emotional response in the first phase was coded and analyzed continuously from video recordings using the software package Interact from Mangold. Coding of phase 1 started as soon as the E started the digital timer and ended as soon as the subject had retrieved the cricket from the E's hand.

### *Dependent measures during phase 1*

In order to assess the extent of emotional reactivity in the subjects, we used several visible and audible behavioral indicators of emotional arousal usually shown by marmosets in stressful or challenging situations. We expected these indicators of emotional reactivity to be more pronounced with the unfamiliar than the familiar experimenters.

### *Behavioral indicators of arousal*

#### 1. Tail-brush

In common marmosets, piloerection is most clearly visible in the brushing of the tail. A brushed tail indicates general vegetative arousal, but not necessarily of negative valence (Dettling et al. 2002).

## 2. “Tsik” vocalizations

Tsik vocalizations are mobbing calls that marmosets typically emit in response to unfamiliar humans, stressful noises, other marmoset groups and potential predators (Martins Bezerra and Souto 2008).

## 3. “Egg” vocalizations

Egg vocalizations have been described as vigilance calls and are usually uttered if an unfamiliar human is approaching or if an individual marmoset finds itself in an exposed position (Martins Bezerra and Souto 2008).

## 4. Self-scratching

Self-scratching has been described as a stress-associated behavior for primates, including marmosets (Bassett et al. 2003).

## 5. Escape

An escape involved the subject rattling and gnawing on the closed door (a gray plastic panel) on top of the experimental compartment in an attempt to leave the test situation.

## 6. Not taking offered food

Not taking offered food was a subject’s refusal to retrieve the cricket from the E’s hand who offered it as a reward in the end of phase 1. From previous experiments, all subjects were used to retrieve rewards from an E’s hand.

The definitions of the dependent variables are summarized in Table 2. The variables tail-brush (1) and escape (5) were measured as states. The beginning and end of a tail-brush state were determined by visible brushing or de-brushing of the tail, captured as a quick change (within a few seconds) in the degree of its piloerection. The duration of an escape state was delimited by the visible and audible start and end of the rattling and gnawing on the door at the top of the experimental compartment. The variables tsik calls (2), egg calls (3), self-scratching (4) and not taking offered food (6) were treated as events. Every individual egg and tsik vocalization was coded as an event. Self-scratching was defined as an event of the subject using a hand or foot to scratch a part of its own body. Not taking the offered food was recorded as a single occurrence of the subject not retrieving the offered food from E’s hand. The durations for the two state variables were measured in seconds and calculated as percentages of the total duration of phase 1. Events were calculated as relative frequencies by dividing the absolute frequency of the behavior by the total duration of phase 1.

## Dependent measures during phase 2

Task participation was defined as the total number of trials the subjects completed in the visible and invisible displacement test.

The subjects’ performance in the cognitive test was live-coded by the E who noted each of the subject’s choices as correct, wrong or non-choice and later verified these choices based on the video clips. Furthermore, we coded from the video clips for each trial in which a subject participated, whether the subject had actually seen the baiting and displacement or not, based on its head and body orientation during the manipulation of the experimenter. If the subject’s head and eyes were oriented in a way that the E’s hand and the containers were in its line of sight at the moment of the baiting and the transposition, a trial was coded as seen by the subject.

Task performance was calculated separately for each test and defined as the number of correct choices in all trials in which the subjects had paid attention to the presentation and therefore actually seen where the food was hidden or displaced.

## Reliabilities and statistical data analysis

In order to assess inter-rater reliability of behavioral coding for phase 1, and of the task performance and attention in phase 2, 20 % of the 56 video clips were coded by a second rater. Krippendorff’s alpha coefficient was calculated for each dependent variable (Hayes and Krippendorff 2007; Krippendorff 2011). For the emotional reactivity variables from phase 1, alpha was generally high: egg calls  $\alpha(12) = 0.99$ , tail-brush  $\alpha(12) = 0.91$ , tsik calls  $\alpha(12) = 0.97$  and escape  $\alpha(12) = 0.99$ . For phase 2, the cognitive task, all choices and non-choices were double-checked based on the video recordings. Correspondence between initially noted choices and double-checked choices was a 100 %. Krippendorff’s alpha for the subjects’ attention in the first object permanence test was  $\alpha(72) = 1.00$  for both the first and second object permanence tasks.

The dependent variables from phase 1 were first subjected to a principal component analysis (PCA without rotation) that established whether they represented the same or different dimensions of the subjects’ emotional reactivity. However, *self-scratching* and *not taking offered food* were not used in this PCA owing to floor effects. Self-scratching was an extremely rare behavior, and with the exception of two male subjects who refused to accept the food from one unfamiliar experimenter, all subjects accepted the offered reward from all four experimenters.

The principal components extracted from the PCA were used as explanatory factors in subsequent general linear

**Table 2** Definitions of the emotional reactivity measures

	Dependent variable	Measurement level	Coded behavior
1	Tail-brush	State/duration	Visible change in piloerection of the tail
2	“Tsik” vocalizations	Event/frequency	Every individual call of a sequence was counted as a vocalization
3	“Egg” vocalizations	Event/frequency	Every individual call of a sequence was counted as a vocalization
4	<i>Self-scratching</i>	<i>Event/frequency</i>	<i>Using a hand or leg to scratch a part of its own body</i>
5	Escape	State/duration	Rattling and gnawing on the door at the top of the experimental compartment
6	<i>Not taking offered food</i>	<i>Single occurrence yes/no</i>	<i>No attempt to manually retrieve the offered cricket from the experimenter’s hand</i>

Four dependent variables (1, 2, 3 and 5) were included in the statistical analysis, while the other two dependent variables (4 and 6 *in italics*) were excluded owing to floor effects

mixed models (GLMMs) with task participation, task performance and attention, as response variables, using the Standard Least Squares option and the restricted maximum likelihood method (REML). In all models, experimenter nested in experimenter group and subject were included as random effects. All statistical analyses were conducted using the software Jmp 10. We used a significance level of  $\alpha = .05$  for all statistical tests, and all tests were two-tailed.

*Is there individual variation in emotional reactivity toward different experimenters?* In order to answer the first question of whether the extent of the subjects’ emotional reactivity differed when confronted with familiar versus unfamiliar experimenters, a separate GLMM was conducted for each of the two components resulting from the PCA (i.e., arousal and avoidance, see below).

*Do individual differences in emotional reactivity affect the subjects’ participation in the cognitive task?* To address the second question of whether elevated emotional reactivity (phase 1) influenced a subject’s motivation to participate in the cognitive task (phase 2), we ran GLMMs for the whole object permanence task with the relative number of completed trials out of twelve as response variable. As fixed effects, we included arousal (as measured by factor 1 of the PCA), avoidance (as measured by factor 2), subject’s sex, and their univariate and bivariate combinations with and without interactions. The best model was identified based on the lowest value of the Akaike information criterion, controlled for small sample sizes (AICc, Hurvich and Tsai 1989). This model selection approach was chosen to account for the limited number of subjects.

Furthermore, we tested whether the extent of a subject’s emotional reactivity affected how attentive it was during the baiting procedure in the first object permanence test and during the baiting and the transposition procedure of the cups in the second test. Therefore, we ran a second series of nine GLMMs with the same fixed and random effects as above, but the response variable was the number of trials a subject had attended to divided by all participated trials.

*Do individual differences in emotional reactivity affect the subjects’ performance in the cognitive task?* We analyzed the third question of whether high emotional reactivity toward the experimenter and the test situation influenced the subjects’ performance in the cognitive task by running several GLMMs separately for each of the two object permanence tests.

Analogous to question 2, we ran two series of nine different GLMMs all of which included the fixed effects subject’s sex, arousal, avoidance and their interactions. Furthermore, we used two different response variables to elucidate how the subjects’ attention to the test procedure influenced their performance, i.e., the total number of correct trials divided by the number of participated trials and the total number of correct trials a subject had attended to divided by all attended trials.

## Results

### Two independent dimensions of emotional reactivity

The PCA on the four emotional reactivity measures in phase 1 (Table 2) revealed two independent dimensions with eigenvalues greater than 1 (2.167 and 1.186). Three dependent variables loaded highly on the first factor that explained 54.2 % of the total variance in the dependent measures: egg calls (0.857), tail-brush (0.840) and tsik calls (0.837). Since all three variables represented audible and visible behavioral signs of arousal, the first factor was named “arousal.” The fourth variable, escape, loaded very highly (0.954) on the second factor which explained 29.7 % of the total variance and was named “avoidance” (see Table 3). To avoid collinearity issues, the two emotional reactivity factors, arousal (factor 1) and avoidance (factor 2), were used for all further statistical analyses.

### Individual variation in emotional reactivity

A first GLMM with the response variable arousal and the fixed effects experimenter group and subject’s sex, and



their interaction, showed no significant main effect of experimenter group on subjects' arousal [ $F(1,2) = 12.88$ ,  $p = .070$ ], although there was a trend for the subjects' arousal to be more pronounced when confronted with unfamiliar as opposed to familiar experimenters. There was, however, a significant effect of the subjects' sex  $F(1,12) = 7.28$ ,  $p = .019$ . A post hoc Student's  $t$  test revealed that male marmosets showed significantly higher arousal than females  $t(12) = 2.70$ ,  $p = .019$  (Fig. 2). Furthermore, the interaction between experimenter group and subjects' sex was highly significant  $F(1,38) = 8.55$ ,  $p < .006$ . A Tukey HSD post hoc test revealed that male marmosets showed significantly more emotional arousal when confronted with unfamiliar experimenters (least square mean = 1.56, SE = 0.41) than with familiar ones (LS mean =  $-0.22$ , SE = 0.41) and than female subjects with both unfamiliar (LS mean =  $-0.44$ , SE = 0.41) and familiar experimenters (LS mean =  $-0.90$ ; SE = 0.41).

A second GLMM with avoidance as response variable revealed a strong trend in the same direction. Again, there was no significant effect of experimenter group  $F(1,2) = 0.31$ ,  $p = .634$  but a significant effect of the subjects' sex  $F(1,12) = 4.86$ ,  $p = .048$ . Male marmosets generally showed more avoidance behavior than females  $t(12) = 2.21$ ,  $p = .048$  (Fig. 2). However, although there was a trend for more pronounced escape behavior of male marmosets in response to the unfamiliar experimenter group, the interaction between experimenter group and subject's sex was not significant  $F(1,38) = 1.42$ ,  $p = .241$ .

Emotional reactivity influences the subjects' participation

#### Participation in the object permanence task

The best model to explain the effect of the subjects' emotional reactivity on their overall participation in the object permanence task (both displacement tests) was the

**Table 3** Two-factor structure of the subjects' emotional reactivity: arousal and avoidance behavior

Dependent variable	F1 arousal	F2 avoidance
Egg calls	<b>0.857</b>	$-0.287$
Tail-brush	<b>0.840</b>	0.360
Tsik calls	<b>0.837</b>	$-0.251$
Escape attempts	0.161	<b>0.954</b>

Loadings of the four dependent variables on the two extracted principal components (factors). Both stress-related vocalizations and piloerection of the tail (tail-brush) loaded highly on the first factor, arousal, which explained 54.2 % of the total variation. The subjects' attempts to leave the situation loaded highly on the second factor, avoidance, which explained a further 29.7 % of the total variation

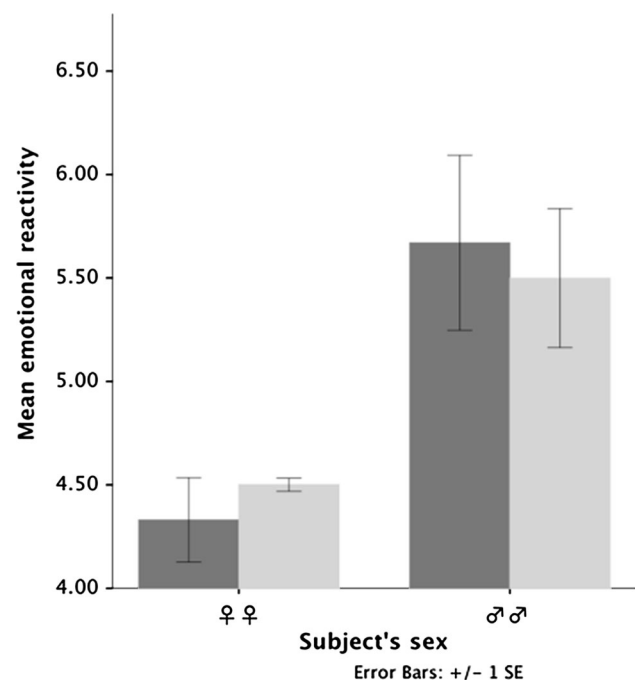
High factor loadings are indicated in bold

one that included subject's sex as the only fixed effect (model 1, see *table S1* in the supplementary material). In this model, a subject's sex had a highly significant effect on the number of completed trials  $F(1,12) = 41.16$ ,  $p < .0001$ . However, the model that additionally included arousal showed nearly identical performance (model 4,  $\Delta AICc = 1.840$ , see *table S1*). In this model, both effects, a subject's sex [ $F(1,13.2)$ ,  $p < .002$ ] and arousal [ $F(1,36.6)$   $p < .024$ ], significantly influenced participation in the test. A Student's  $t$  test [ $t(12) = 6.42$ ,  $p < .0001$ ] revealed that male subjects completed significantly fewer trials (LS Mean = 0.46, SE = 0.09) than females (LS Mean = 0.99, SE = 0.09), who in fact all participated in all six trials of the visible displacement test and in most trials of the invisible displacement test (Fig. 3a).

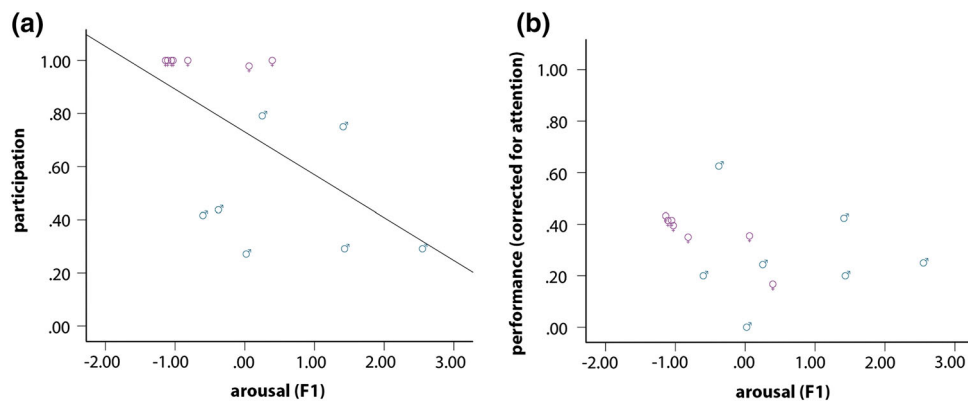
The relatively small difference between the two models ( $\Delta AICc = 1.836$ ) indicates that both variables, a subject's sex and its arousal, explain the individual variance in participation equally well. Their interaction was not significant (model 5, *table S1*).

#### Attention in the object permanence task

The best model for the response variable attention, the number of seen trials out of all participated trials, was the

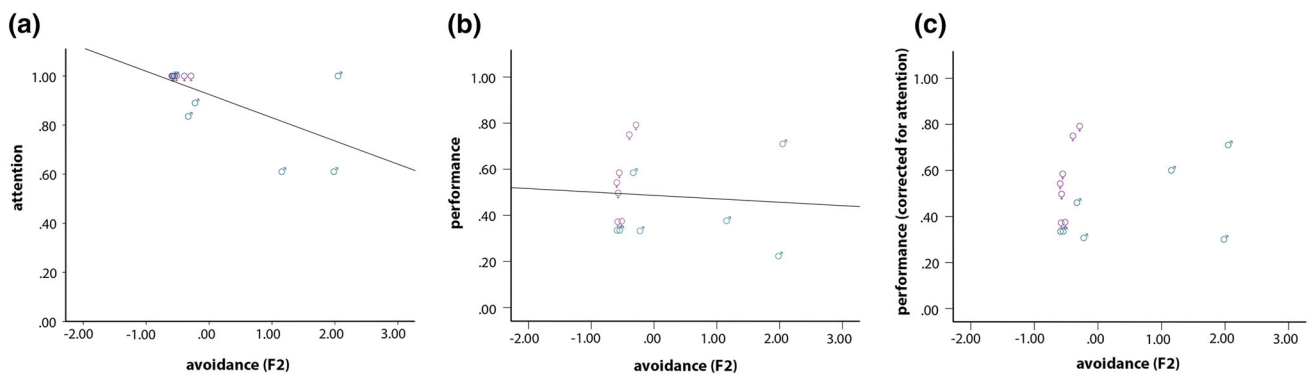


**Fig. 2** Effect of a subject's sex on its emotional reactivity. The dark gray bars represent the subjects' mean arousal (as measured by factor 1 of the PCA), the light gray bars their mean avoidance levels (as measured by factor 2). The two left bars represent female subjects (♀♀), the two right bars males (♂♂). Note: To better visualize the values that resulted from the PCA, they were transformed by adding a factor of 5 so that all values appeared greater than zero



**Fig. 3** A subject's sex and emotional arousal affected its participation but not its performance in the object permanence task. Mean percentages of (a) participated trials of a subject in the object

permanence task, and (b) correct trials in the second test, invisible displacement, corrected for attention. A subject's sex is indicated by its gender symbol (♀ = female, ♂ = male)



**Fig. 4** Emotional avoidance affected a subject's attention but not its performance in the first object permanence test, visible displacement. Shown is the effect of a subject's avoidance levels on (a) the mean percentage of trials it paid attention to, (b) its uncorrected

performance: the mean percentage of correct of all participated trials, and (c) its corrected performance: the mean percentage of correct trials it paid attention to in the visible displacement test. A subject's sex is indicated by its gender symbol (♀ = female, ♂ = males)

one which only included the fixed effect arousal  $F(1,31.3) = 30.94$ ,  $p < .0001$  (model 2,  $AICc = 40.780$ , see *table S2* in the supplementary material). Highly aroused subjects saw the test procedure in fewer participated trials than subjects with low arousal. Moreover, although there was no significant effect of avoidance on overall attention, avoidant subjects were less attentive in the first object permanence test  $F(1,21.54) = 9.16$ ,  $p = .006$  (Fig. 4a).

Emotional reactivity does not influence cognitive test performance

#### Performance in test 1 visible displacement

When using the subjects' *uncorrected performance* as response variable, i.e., the number of correct trials of all participated trials in the first test, the best of the nine models based on the lowest Akaike information criterion was the one that included avoidance as the only fixed effect  $F(1,38.1) = 7.29$ ,  $p = .010$  (model 3,  $AICc = 18.035$ , see

*table S3a* in the supplementary material). Highly avoidant subjects, who were predominantly males, made fewer correct choices than less avoidant ones (Fig. 4b).

When using *corrected performance* as response variable, i.e., the number of correct trials in which the subject saw the baiting of the cup, the best model was still the one with the fixed effect avoidance  $F(1,36.01) = 2.40$ ,  $p = .130$  (model 3,  $AICc = 20.090$ , see *table S3b*) but a subject's avoidance level no longer had a significant effect on its performance (Fig. 4c). Moreover, the model that only included subject's sex as fixed effect (model 1,  $\Delta AICc = 0.315$  see *table S3b*) and the model including arousal only (model 2,  $\Delta AICc = 1.876$ , see *table S3b*) can be considered as equivalent based on their  $\Delta AICc$  values lower than 2. None of these effects were significant.

#### Performance in test 2 invisible displacement

When we used the *uncorrected performance* as response variable, the best model included only subject's sex (model

1,  $AICc = -4.251$ , see table S4 in the supplementary material). Male subjects tended to perform more poorly than females, but this effect was not significant  $F(1,13.85) = 1.90$ ,  $p = .189$ .

When we used the *corrected performance* as response variable and thus controlled for attention, the best model was, as for test 1, the one including avoidance as single fixed effect  $F(1,18.95) = 1.05$ ,  $p = .318$  (model 3, see table S4). However, the two models that only included sex (model 1,  $\Delta AIC = 1.319$ ) and arousal (model 2,  $\Delta AIC = 0.368$ ) resembled it closely (see table S4). None of these effects were significant (Fig. 3b).

## Discussion

The aim of this study was to quantitatively assess whether and how a subject's emotional reactivity toward an experimenter and the test situation impacts participation and performance in a cognitive task.

In a two-phase experiment conducted in a single session, we quantified the extent of emotional reactivity of common marmosets toward four different experimenters. We then assessed whether and how emotional reactivity affected the subjects' participation and performance in a subsequent object permanence task. To maximize variation in the subjects' emotional reactivity, we confronted subjects with two groups of experimenters whom we expected to elicit high variation in emotional reactivity, two male experimenters, who were complete strangers to the subjects, versus two familiar female experimenters. In the first phase of the experiment, the experimenters performed an identical standardized sequence of behavioral actions in front of the subjects. We found considerable variability in the subjects' spontaneous emotional response as measured by four behavioral indicators that represented two independent dimensions of emotional reactivity, arousal and avoidance. The marmosets behaviorally expressed high arousal by showing piloerection of the tail and emitting two distress-related vocalizations, whereas avoidance was expressed by attempts to leave the experimental situation. Overall and in accordance with our expectation, male subjects were more emotionally aroused than females when confronted with the experimenters, in particular with unfamiliar ones. They also tended to try to avoid the situation more often than female marmosets, independently from experimenter familiarity.

The extent of emotional reactivity during phase 1 had affected the male subjects' participation in the subsequent object permanence task in phase 2, whereas the participation of females was largely unaffected. Moreover, marmosets who were more emotionally reactive in phase 1 tended to be less attentive to the test procedure when they eventually participated in the cognitive test.

In contrast to their participation, the subjects' cognitive performance in the object permanence task was not affected by emotional reactivity, given we controlled their performance for whether they had paid attention to the experimental procedure or not. However, when not controlling for attention, individuals who had shown more attempts to leave the situation (avoidance) in phase 1 made more mistakes. Together, the results suggest that individuals who were more emotionally reactive in phase 1 were less likely to be attentive to the task prior to making their choice and made more mistakes. Thus, controlling for attention is vital in particular for individuals who show higher emotional reactivity. We conclude that in common marmosets, emotional reactivity toward an experimenter affects participation and attention but not performance in cognitive tasks. An interesting direction for future studies is to disentangle what factors are responsible for the higher emotional reactivity of some individuals. Besides a subject's sex, possible factors are its personality, its familiarity with the experimenter, the experimenter's gender, similar personalities of both the subject and the experimenter, or a combination of these factors.

In our study, male marmosets showed a stronger emotional response, were less motivated to participate, completed fewer trials and were less likely to attend to the presentation and experimental manipulation of the cups and the reward. However, in the trials in which they did participate and paid attention to the presentation and manipulation of these stimuli, their cognitive performance did not differ from their female conspecifics' performance. This result contrasts with the idea that female common marmosets possess better cognitive abilities than males (Yamamoto et al. 2004; Brown et al. 2010). Our results suggest that these reported sex differences in cognitive performance in marmosets are best understood in terms of motivational and attentional sex differences, such as a higher interest of female marmosets to work for food and of male marmosets to remain vigilant to the environment (see also Koenig 1998), rather than reflecting a true sex difference in their cognitive abilities.

Our results have two implications for conducting experimental tests with marmosets, and arguably non-human primates in general. First, comparable results may be achieved with different experimenters, even if subjects vary individually in how they behaviorally respond to these different experimenters. It is crucial, however, to control for whether the subjects are attentive to the presentation and manipulation of the experimental stimuli, because particularly highly reactive individuals may participate in the task without paying attention to it, perhaps in order to quickly finish the experiment. Ideally, the experimenter should obtain the subject's full attention before starting a test trial and closely attend to whether the subject stays

attentive during the whole test procedure. However, monitoring a subject's attention during testing can be difficult, because the experimenter has to avoid to inadvertently give visual cues to the subject and cannot look at it during the actual test trial. Therefore, it is vital to determine from video recordings for every trial whether the subject had been fully attentive or not, and to correct its performance accordingly.

The second implication is that excluding highly reactive animals from testing does not necessarily bias the results. In fact, the alternative explanation of the trait anxiety effect in macaques (Toxopeus et al. 2005), i.e., that the adverse effect of trait anxiety on cognitive performance results from attentional biases rather than cognitive ability per se, is supported by the present set of findings. Likewise, our results are consistent with the ones obtained by Morton et al. (2013) who basically showed a participation effect rather than a truly cognitive one.

Based on our findings, we suggest that differences in emotional reactivity do not necessarily have an impact on cognitive performance. Thus, it is justifiable both to use different experimenters and to exclude very highly aroused and unmotivated subjects from testing. However, observable behavioral indicators of elevated emotional arousal should sensitize the experimenter to evaluate an individual subject's attentional state prior to and during cognitive testing.

**Acknowledgments** We thank Prof. Carel van Schaik for discussion, Dr. Karin Isler and Christa Finkenwirth for statistical advice, Caroline Jordan for the graphic implementation of Fig. 1, and our animal keepers, Heinz Galli and Thomas Bischof, for help with the preparation of the experimental room prior to this study. We also thank two anonymous reviewers for their helpful comments on the initial manuscript. The study was funded by the Swiss National Science Foundation (project number: 310030\_130383). It was performed under guidelines established by the National Veterinary Office of Switzerland and licensed to be performed by the Veterinary Office of the Canton of Zürich. The authors declare they have no conflict of interest.

## References

- Bassett L, Buchanan-Smith HM, McKinley J, Smith TE (2003) Effects of training on stress-related behavior of the common marmoset (*Callithrix jacchus*) in relation to coping with routine husbandry procedures. *J Appl Anim Welf Sci* 6(3):221–233
- Brown J, Kaplan G, Rogers LJ, Vallortigara G (2010) Perception of biological motion in common marmosets (*Callithrix jacchus*): by females only. *Anim Cogn* 13(3):555–564
- Burkart JM, Fehr E, Efferson C, van Schaik CP (2007) Other-regarding preferences in a nonhuman primate: common marmosets provision food altruistically. *PNAS* 104(50):19762–19766
- Burkart JM, Strasser A, Foglia M (2009) Trade-offs between social learning and individual innovativeness in common marmosets, *Callithrix jacchus*. *Anim Behav* 77(5):1291–1301
- Crofoot MC, Lambert TD, Kays R, Wikelski MC (2010) Does watching a monkey change its behaviour? Quantifying observer effects in habituated wild primates using automated radiotelemetry. *Anim Behav* 80(3):475–480
- Dettling AC, Feldon D, Pryce CR (2002) Repeated parental deprivation in the infant common marmoset (*Callithrix jacchus*, primates) and analysis of its effects on early development. *Biol Psychiatry* 52:1037–1046. doi:10.3758/BRM.41.2.396
- Ford SM, Porter LM, Davis LC (2009) The smallest anthropoids: the marmoset/callimico radiation (developments in primatology: progress and prospects). Springer, New York
- Hayes AF, Krippendorff K (2007) Answering the call for a standard reliability measure for coding data. *Commun Methods Meas* 1(1):77–89
- Hurvich CM, Tsai CL (1989) Regression and time series model selection in small samples. *Biometrika* 76:297–307
- Koenig A (1998) Visual scanning by common marmosets (*Callithrix jacchus*): functional aspects and the special role of adult males. *Primates* 39(1):85–90
- Krippendorff K (2011) Computing Krippendorff's alpha-reliability. Retrieved from [http://repository.upenn.edu/asc\\_papers/43](http://repository.upenn.edu/asc_papers/43)
- Martins Bezerra B, Souto A (2008) Structure and usage of the vocal repertoire of *Callithrix jacchus*. *Int J Primatol* 29:671–701
- Mendes N, Huber L (2004) Object permanence in common marmosets (*Callithrix jacchus*). *J Comp Psychol* 118:103–112
- Morton FB, Lee PC, Buchanan-Smith HM (2013) Taking personality selection bias seriously in animal cognition research: a case study in capuchin monkeys (*Sapajus apella*). *Anim Cogn* 16:677–684. doi:10.007/s10071-013-0603-5
- Rosenthal R (1963) On the social psychology of the psychological experiment: the experimenter's hypothesis as unintended determinant of experimental results. *Am Sci* 51(2):268–283
- Schellinck HM, Cyr DP, Bron RE (2010) How many ways can mouse behavioral experiments go wrong? Confounding variables in mouse models of neurodegenerative diseases and how to control them. *Advances in the Study of Behavior* 41:255–366
- Siegwarth N, Larkin KT, Kemmner C (2012) Experimental effects on cardiovascular reactivity and task performance during mental stress testing. *Psychol Record* 62(1):69–81
- Sorge RE, Martin LJ, Isbester KA, Sotocinal SG et al (2014) Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nat Methods* 11:629–632. doi:10.1038/nmeth.2935
- Toxopeus IB, Sterck EHM, van Hooft JARAM, Spruijt BM, Heeren TJ (2005) Effects of trait anxiety on performance of socially housed monkeys in a learning test. *Behaviour* 142(9–10):1269–1287. doi:10.1163/156853905774539373
- Yamamoto ME, Domeniconi C, Box H (2004) Sex differences in common Marmosets (*Callithrix jacchus*) in response to an unfamiliar food tasks. *Primates* 45:249–254
- Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. *J Comp Neurol Psychol* 18:459–482. doi:10.1002/cne.920180503