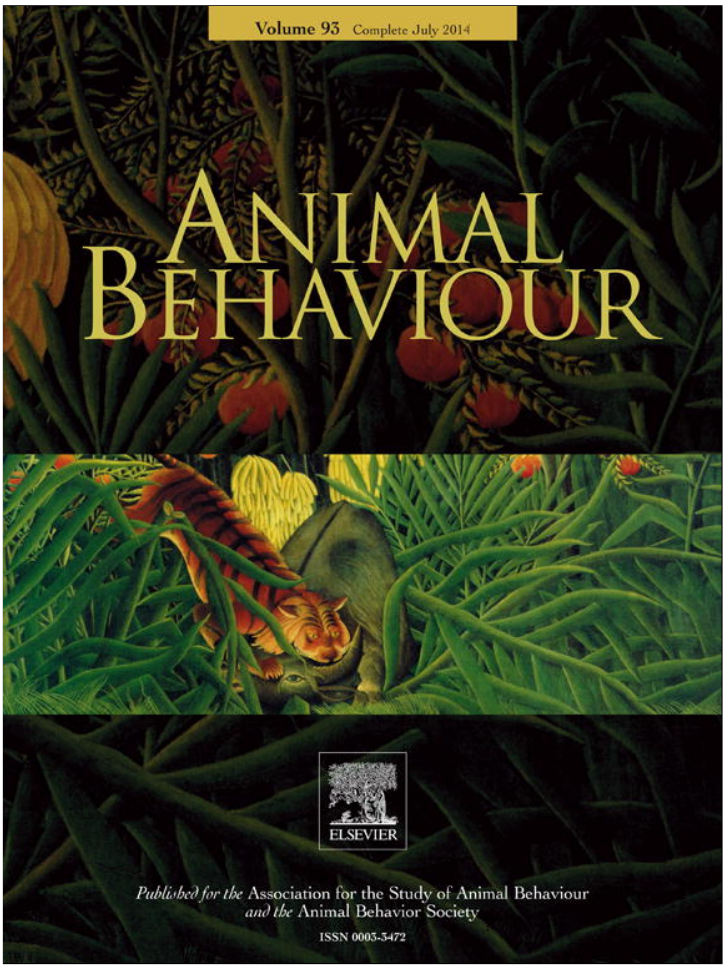
Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/authorsrights

Author's personal copy

Animal Behaviour 93 (2014) 49–58

Contents lists available at ScienceDirect

Animal Behaviour

journal homepage: www.elsevier.com/locate/anbehav

Competition-induced stress does not explain deceptive alarm calling in tufted capuchin monkeys

Brandon C. Wheeler [a,b,*], Barbara Tiddi [a,b], Michael Heistermann [c]

[a] Cognitive Ethology Laboratory, German Primate Center, Germany
[b] Courant Research Centre Evolution of Social Behaviour, University of Göttingen, Germany
[c] Endocrinology Laboratory, German Primate Center, Germany

ARTICLE INFO

Article history:
Received 2 December 2013
Initial acceptance 6 January 2014
Final acceptance 18 March 2014
Available online 21 May 2014
MS. number: 13-00999R

Keywords:
alarm call
Cebus apella nigritus
cognition
communication
cortisol
Machiavellian intelligence
New World primate
tactical deception
within-group contest

Tactical deception has long attracted interest because it is often assumed to entail complex cognitive mechanisms. However, systematic evidence of tactical deception is rare and no study has attempted to determine whether such behaviours may be underpinned by relatively simple mechanisms. This study examined whether deceptive alarm calling among wild tufted capuchin monkeys, *Cebus apella nigritus*, feeding on contestable food resources can be potentially explained by a physiological mechanism, namely increased activation in the adrenocortex and the resulting production of glucocorticoids (GCs; 'stress hormones'). This was tested experimentally in Iguazú National Park, Argentina, by manipulating the potential for contest competition over food and noninvasively monitoring GC production through analysis of faecal hormone metabolites. If deceptive false alarms are indeed associated with adrenocortical activity, it was predicted that the patterns of production of these calls would match the patterns of GC output, generally being higher in callers than noncallers in cases in which food is most contestable, and specifically being higher in callers on those occasions when a deceptive false alarm was produced. This hypothesis was not supported, as (1) GC output was significantly lower in association with the experimental introduction of contestable resources than in natural contexts wherein the potential for contest is lower, (2) within experimental contexts, there was a nonsignificant tendency for noncallers to show higher GC output than callers when food was most contestable, and (3) individuals did not show higher GC levels in cases in which they produced deceptive alarms relative to cases in which they did not. A learned association between the production of alarms and increased access to food may be the most likely cognitive explanation for this case of tactical deception, although unexplored physiological mechanisms also remain possible.
© 2014 The Authors. Published on behalf of The Association for the Study of Animal Behaviour by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

The well-documented relationship between social group size and the size of the brain generally and the neocortex specifically has lent widespread support for the so-called social brain hypothesis (Dunbar & Shultz, 2007; Dunbar, 1998; but see Barton, 2012). While there are a number of reasons why individuals with larger brains may be favoured in highly social environments, the Machiavellian intelligence hypothesis (MIH) posits that larger brains are advantageous for social animals specifically because increased cognitive abilities allow individuals to outwit their groupmates in the competitive interactions that are a near-universal result of group living (Byrne & Whiten, 1988). In particular, the MIH predicts tactical deception, that is 'acts from the normal repertoire of the agent, deployed such that another individual is likely to misinterpret what the acts signify, to the advantage of the agent' (Byrne & Whiten, 1990, p. 3), to be common among large-brained, social taxa, especially primates.

In support of the MIH, there have been many anecdotal observations of apparent tactical deception in a wide range of anthropoid primates (Byrne & Whiten, 1990), with more such anecdotal observations reported for taxa with a larger neocortex ratio (Byrne & Corp, 2004). Unfortunately, the anecdotal nature of these observations has hampered systematic investigation into the proximate mechanisms underpinning the behaviours. It is thus unclear whether these observations of tactical deception are examples of flexibly deployed behaviours underpinned by an intention to

* Correspondence: B. Wheeler, Cognitive Ethology Laboratory, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany.
E-mail address: bcwheeler43@gmail.com (B. C. Wheeler).

http://dx.doi.org/10.1016/j.anbehav.2014.04.016
0003-3472/© 2014 The Authors. Published on behalf of The Association for the Study of Animal Behaviour by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

Author's personal copy

change the targets' behaviour or beliefs (and thus involving, respectively, first- or second-order intentionality; see Shettleworth, 2010) as the MIH would predict. Alternatively, such apparently complex behaviours may involve zero-order intentionality (i.e. nonintentional; see Shettleworth, 2010) and be better explained as rather inflexible and innate responses to external stimuli mediated, for example, by hormonal states (e.g. see Bshary et al., 2011).

Systematic examination of the proximate mechanisms underlying tactical deception has been complicated by the rarity of such behaviours, a consequence of the fact that tactical deception deployed too frequently is likely to be ignored, reducing its effectiveness (Johnstone & Grafen, 1993; Maynard Smith & Harper, 2003). Antipredator communication systems, however, are vulnerable to high rates of functionally deceptive signalling because the cost of ignoring a signal that honestly indicates the presence of a predator is potentially death, outweighing the costs of responding to deceptive false alarms (Searcy & Nowicki, 2005). Indeed, an increasing number of studies across a range of taxa, including passerine birds, ungulates, sciurid rodents and primates, have provided systematic evidence that false alarm signals given in competitive feeding and mating contexts function to provide the signaller with a tactical advantage by eliciting unnecessary antipredator behaviours in receivers (Bro-Jørgensen & Pangle, 2010; Flower, 2011; Møller, 1988; Munn, 1986; Tamura, 1995; Wheeler, 2009). However, while these studies show that functionally deceptive antipredator signals can occur frequently enough to be examined systematically, no study to date has attempted to directly examine the underlying proximate mechanisms.

Tufted capuchin monkeys, *Cebus apella nigritus* (synonymous with *Sapajus nigritus*), have been shown to produce terrestrial predator-associated alarm calls ('hiccups') in competitive feeding contexts, but in the absence of predators, in a manner that is consistent with an interpretation of tactical deception (Wheeler, 2009). Specifically, these alarm calls, which are acoustically indistinguishable from alarms elicited by predatory stimuli (Wheeler & Hammerschmidt, 2013), are given far more often in experimental feeding contexts wherein high-value foods are presented in discrete, contestable patches than they are in natural contexts (Wheeler, 2010a). The interpretation of these as tactical deception is supported by the fact that listeners commonly run higher into the canopy and out of the food patch, while callers do not show such predator evasion behaviours, but instead move into the food patch (Wheeler, 2009). Furthermore, these spontaneously produced alarm calls tend to be given by relatively low-ranking individuals sitting in the immediate vicinity of the food patch, and they occur most often when food is highly clumped, and therefore easily contested and monopolized by dominants (Janson, 1996; Wheeler, 2009). These systematic observations indicate that false alarm calls function to deceive more dominant competitors, and suggest this as an ideal system to examine the proximate mechanisms underpinning the behaviour.

While at least superficially providing support for the MIH, functionally deceptive alarm calling among capuchins might be better explained as a relatively inflexible behaviour mediated by emotional mechanisms and their hormonal correlates than by the more cognitively complex mechanisms (e.g. reasoning or learning) necessary for first- or second-order intentionality. Such a 'simple' explanation in fact appears likely in this case given that vocal production in nonhuman primates (and most other terrestrial mammals) is generally rather inflexible in terms of the contexts in which a given call can be produced, resulting from an innate relationship between a particular call type and underlying internal states (e.g. Hammerschmidt & Fischer, 2008; Owren, Dieter, Seyfarth, & Cheney, 1992; Seyfarth & Cheney, 2010; Wheeler & Fischer, 2012). Encounters with predators are known to cause activation of the adrenocortex and an associated increase in the production of glucocorticoids (GCs; 'stress hormones') in many prey taxa (e.g. Arlet & Isbell, 2009; Clinchy, Sheriff, & Zanette, 2013; Cockrem & Silverin, 2002; Mateo, 2010; Monclús, Rödel, Palme, Von Holst, & de Miguel, 2006). Likewise, several studies of primates and other mammals, including capuchins, have demonstrated a positive relationship between circulating GC levels and rates of production of vocalizations associated with terrestrial predators (Blumstein, Patton, & Saltzman, 2006; Boinski, Gross, & Davis, 1999; Cross & Rogers, 2006; see also Fichtel & Kappeler, 2002; Mateo, 2010), with one such study suggesting a causal relationship in macaques (Bercovitch, Hauser, & Jones, 1995). Furthermore, a study of captive bonobos, *Pan paniscus*, has shown that adrenocortical activity can also be affected by food distribution, with higher GC levels seen when food is clumped relative to when it is dispersed (Hohmann, Mundry, & Deschner, 2009). Given that the intensity of contest competition (i.e. dominance rank-based skew in energy gain) is well documented to increase with food clumpiness in social foragers (reviewed in Koenig, 2002; Wheeler, Scarry, & Koenig, 2013), competition for clumped foods might be expected to elicit a stronger stress response in relatively lower-ranking individuals (Foerster & Monfort, 2010; see also Abbott et al., 2003). It is thus plausible that the use of deceptive false alarms in capuchins and other animals (e.g. Bro-Jørgensen & Pangle, 2010; Møller, 1988; Munn, 1986) results from an underlying relationship between GCs and the propensity to produce predator-associated calls.

This study aimed to test whether variation in GC production indeed provides a plausible proximate explanation for the documented deceptive alarm-calling behaviour of tufted capuchin monkeys. If this is the case, then the patterns of GC production should match the patterns of false alarm production, with individuals that produce more false alarm calls in particular contexts showing higher GC levels in those contexts than individuals that produce fewer deceptive calls. We thus predicted that, relative to individuals who do not produce deceptive false alarms, individuals with a higher propensity to give deceptive calls should show higher GC output in association with (1) experimental contexts in which resources are presented in contestable patches than in natural contexts wherein the potential for contest competition is reduced, and (2) experimental contexts in which food is highly clumped (and contestable) than in experimental contexts in which food is relatively dispersed (and therefore less contestable). Last, we predicted that (3) on those specific occasions in which individuals produced deceptive false alarms, GC levels would be higher than on occasions in which no alarms were given. Although support for these predictions would not necessarily indicate a causal relationship between the production of GCs and deceptive false alarms, nor rule out the possibility that the production of false alarms is a result of physiological stress working in conjunction with cognitive mechanisms (see Mateo, 2008; Soares et al., 2010), a lack of support would in effect rule out the possibility that high GC levels are a necessary precondition for the production of these calls. Direct tests of the relationship between GC levels and deceptive false alarm call production are thus an important first step in narrowing down the possible proximate explanations for this example of tactical deception.

METHODS

Study Site and Subjects

Data were collected from June to August 2010 and June to August 2011 in Iguazú National Park in northeastern Argentina (25°40′S, 54°30′W). The site is part of the South American Atlantic

Author's personal copy

Forest and is characterized by humid, subtropical rainforest. Seasonal variation in temperature results in low availability of the predominant resources exploited by capuchins (fruits and invertebrates) during the austral winter (Brown & Zunino, 1990). The most common predators of capuchins in Iguazú include ocelots, *Leopardus pardalis*, pumas, *Puma concolor*, and hawk-eagles, *Spizaetus* spp. (Janson, Baldovino, & Di Bitetti, 2012).

Tufted capuchins live in stable social groups of up to 45 individuals (Janson et al., 2012). Within-group dominance relationships are despotic and linear, with dominant individuals winning contests over food and preferred spatial positions (Di Bitetti & Janson, 2001; Janson, 1996). Data for the current study came from two study groups, the Macuco group (26 individuals; studied in 2010) and the Rita group (19 individuals; studied in 2011), in which all individuals were readily recognizable based on natural fur and facial patterns. The current analysis is limited to adults/subadults ($\geq$4.5 years old) and juveniles (ca. 2.5–4.5 years). Further details on the study site and subjects can be found in Janson et al. (2012).

Experimental Protocol

To manipulate food contestability, we conducted controlled provisioning experiments in which the availability and spatial distribution of a high-value food (banana pieces) were varied using ca. 1 m $\times$ 1 m wooden platforms suspended from tree branches at a height of 3–10 m above the ground by a system of ropes and pulleys (see Janson, 1996). The platforms are anchored to the ground to prevent swinging, and the study subjects are well habituated to feeding on these substrates due to nearly two decades of research (Janson, 1996); thus the design of the platforms would not itself be expected to induce stress in the monkeys. Experiments were conducted during the winter months, when capuchins in Iguazú rely heavily on low-value and dispersed fallback foods (Brown & Zunino, 1990), allowing for maximum control over the spatiotemporal distribution, and thus the contestability, of the subjects' preferred food (Janson, 1996). Five experimental sites, separated by approximately 250–350 m, were simultaneously set up within the home range of the study groups. Food contestability was manipulated by varying the number of platforms within each experimental site from one or two platforms per site (clumped condition) to four platforms per site (dispersed condition), with each platform in a site separated from the others by 10–20 m. Similar manipulations of food contestability have previously been shown to affect rank-related skew in energy gain in this population (Janson, 1996). However, the rate of agonistic behaviour (especially displacements) that occurs in association with both the clumped and dispersed conditions used here greatly exceeds that which typically occurs under natural conditions (Wheeler & Tiddi, n.d.), suggesting that even the dispersed condition elicits high levels of contest competition relative to natural contexts.

For the smaller study group, eight average-sized bananas were each cut into approximately six 2–3 cm pieces and were distributed evenly across all the platforms at a site. For the larger study group, 10 bananas per site were normally used, although only six bananas per site were used for a portion of one period in which only a single platform per site was used. We reduced the number of bananas in this case because when we used 10 bananas per site, the group's alpha male tended to monopolize the single platform and become satiated after visiting only one or two sites, resulting in the group subsequently not visiting additional experimental sites.

Although all experiments were conducted during the austral winter, experimental conditions were changed every 10 days throughout the study periods, alternating between periods with and without feeding platforms, in order to eliminate the potentially confounding effect of long-term temporal variation in GC output (e.g. Ostner, Kappeler, & Heistermann, 2008). During 10-day periods in which platforms were used, all five experimental sites were set up according to the same condition (i.e. clumped or dispersed), thus allowing the subjects to receive the same experimental treatment up to five times per day for 10 consecutive days (note, though, that periods of clumped resources varied between pure one-platform set-ups, pure two-platform set-ups and combinations of the two set-ups). The food distribution during a given 10-day period with platforms differed from the previous 10-day period. This resulted in a total of 145 provisioning experiments with the larger study group (clumped condition: 89 during three 10-day periods; dispersed condition: 56 during two 10-day periods) and 182 with the smaller group (clumped condition: 99 during three 10-day periods; dispersed condition: 83 during two 10-day periods). Each group was sampled for four 10-day periods without provisioning.

Platforms with banana pieces were generally raised once the group arrived at the site. In cases in which one or a few group members arrived at a site prior to the rest of the group, platforms were raised as the bulk of the group arrived. Although the smaller study group was highly cohesive during their study period, the larger group was less cohesive, resulting in a number of provisioning experiments in which relatively few individuals arrived at the platform site. Because the number of individuals participating may change the intensity of contest competition (see Wheeler, Scarry, et al., 2013), we considered for our analyses only those experiments in which at least half of the adult and juvenile individuals were present; provisioning experiments with fewer individuals were not considered for any of the analyses. In cases in which a majority arrived for a given provisioning experiment but a subgroup remained behind and did not visit the experimental site, we noted the group members not present. It generally took a group 5–9 min (mean 6.6 min) to remove all banana pieces from the platforms, although this ranged from approximately 1 to 20 min.

Observational Methods

One observer (B.W.) conducted all-occurrence sampling during 278 of the provisioning experiments (approximately 30.6 h of observation) to note the production of all 'high-urgency' terrestrial predator alarms (i.e. bouts consisting of at least two 'hiccup' calls given in quick succession; see Wheeler, 2009; Wheeler, 2010a; Wheeler, 2010b). Such calls were classified as spontaneous false alarms if no eliciting stimulus (including aggression, predators and predator-like stimuli) could be identified and if no additional antipredator behaviours (including escape responses and sudden vigilance beyond the caller's immediate substrate) were performed in conjunction with calling (see Wheeler, 2009; Wheeler, 2010a; Wheeler & Hammerschmidt, 2013). Whenever possible, the identity of the caller was noted. We used the frequency with which individuals were observed to give spontaneous false alarms to assign them to particular alarm-calling classes: 16 individuals who were never identified as producing spontaneous false alarms during feeding platform experiments were categorized as 'noncallers', nine individuals observed to call at least once but in fewer than 1.6% of the platform experiments in which they participated were categorized as 'infrequent callers', and eight individuals observed to call in at least 2.5% of the platform experiments in which they participated were classified as 'frequent callers'. The cutoff between frequent and infrequent callers was based on a natural break in calling rates across individuals.

In addition to observations during provisioning experiments, study groups were followed continuously from the morning to evening sleeping sites every day for the entire duration of their respective study seasons. We noted all observed events considered

Author's personal copy

potentially stressful, including intense aggressive interactions within or between groups, physical injuries and predator encounters (including simulated predators; see Wheeler & Hammerschmidt, 2013). All-occurrence data on the production of bouts of terrestrial predator-associated calls collected by one observer (B.W.) during approximately 593 contact hours were used to measure rates of false alarm calling in natural contexts.

Determination of Glucocorticoid Output

Adrenocortical activity was measured noninvasively through measurement of faecal glucocorticoid metabolites (fGCM). Owing to the long time lag between the circulation of native hormones and the eventual excretion of their metabolites (typically 20–48 h in mammals; see Schwarzenberger, Möstl, Palme, & Bamberg, 1996) and the fact that fGCM levels in a given sample typically represent an average of GC output over a period of hours, it is generally considered difficult at best to study the effects of acute stressors, such as the potential effect of the provisioning experiments, through analysis of fGCM (Anestis, 2010). However, the biology of fGCM excretion in capuchins and the experimental design used in the current study facilitates such analysis. First, peak or near-peak fGCM levels associated with a particular stressor are found in samples excreted 2–4 h after the event, and return to baseline levels within ca. 6–8 h (Wheeler, Tiddi, Kalbitzer, Visalberghi, & Heistermann, 2013); samples collected during the 3 h period that occurred 2–5 h following one or more provisioning experiments can thus be reliably associated with those events. Second, because the experimental design resulted in subjects receiving the same treatment (i.e. no provisioning, clumped food treatment or dispersed food treatment) multiple times per day and for several consecutive days, any time averaging of fGCM levels that may occur in a given sample should still reflect the same experimental condition. Based on these considerations, each sample was coded based on whether or not it was associated with a provisioning experiment (yes/no) in the temporal window 2–5 h prior to defecation. Furthermore, for those samples associated with a provisioning experiment, samples were coded as associated with either the clumped or dispersed condition.

Faecal samples were collected opportunistically from identified adult and juvenile individuals, with the time of defecation noted. Samples were collected in polypropylene tubes within 30 min of defecation and placed in an insulated icepack. Upon return to the field station (within 6 h of sample collection), samples were stored frozen at −15 °C for one to several days until the hormone metabolites were extracted from the wet faeces. Extraction procedures followed the 'field extraction' method described by Wheeler, Tiddi, et al. (2013). Briefly, 0.4–0.6 g of wet, homogenized faeces was extracted with 5 ml of 80% ethanol by vortexing for 5 min; in cases in which 0.3 g of faeces or less could be obtained, only 3 ml of ethanol was used to keep the ethanol to faeces ratio consistent (Palme, Touma, Arias, Dominchin, & Lepschy, 2013). After vortexing, samples were centrifuged at 2000 RPM for 10 min and 1 ml of the supernatant was removed and stored in a 2 ml polypropylene tube wrapped with laboratory film. Following the extraction, the faeces were dried completely and the dry weight determined. Extracts were stored in a refrigerator (for samples collected in 2010) or freezer (for samples collected in 2011) until transported to the laboratory for analysis (during the period of transport, samples were kept at ambient temperatures for 2–3 days, and were subsequently stored in a freezer until analysis). Samples were measured by enzyme immunoassay (see below) within 7–9 months of collection; fGCM concentrations in samples extracted and stored using this method have been shown to be stable for at least 12 months (Wheeler, Tiddi, et al., 2013).

We measured fGCM concentrations in the extracted samples on microtitre plates using a validated corticosterone enzyme immunoassay (Wheeler, Tiddi, et al., 2013) following methods described by Heistermann, Palme, and Ganswindt (2006). Detailed information on assays, including antibodies, standards and labels, and assay sensitivities, can be found in Heistermann et al. (2006). Intra-assay coefficients of variation (CVs) of high- and low-value quality controls were 6.3% and 7.9%, respectively. Corresponding interassay CVs were 10.6% and 11.7%, respectively. All samples from a given individual were measured on a maximum of two microtitre plates (one if fewer than 35 samples for that individual were measured) in order to minimize the contribution of interassay variation to the measured intraindividual variation in GC output. Hormone metabolite concentrations are expressed as ng/g dry faecal weight.

Ethical Note

Although feeding platform experiments induce contest competition and rates of agonistic behaviour (primarily in the form of low-intensity aggression or, most often, spatial displacements without overt aggression) above baseline levels (Di Bitetti & Janson, 2001; Janson, 1996; Wheeler & Tiddi, n.d.), intense aggression at the platforms was rare and we did not observe any cases of aggression at platforms that led to physical injuries in the study subjects. Furthermore, the experiments did not induce increases in physiological stress in the subjects (see Results). All aspects of the study were approved by the Animal Welfare Officer at the German Primate Center and by the Argentine Administration of National Parks (permit no. NEA 142), and adhered to the legal requirements of Argentina.

Data Analysis

Prior to testing the main predictions, we first confirmed their soundness by examining whether contextual variation in the production of false alarms during the current study period matched previous observations, as described above, and on which the predictions were based (i.e. whether false alarm calls occurred more often during provisioning experiments than in natural contexts, and whether, within provisioning contexts, such calls occurred more often in the clumped than dispersed food contexts). We tested this with two Wilcoxon signed-ranks tests that compared the rate of production of spontaneous false alarm calls (i.e. number of times each individual was observed to call/total observation time for each individual in each context) for each individual categorized as either an infrequent or frequent caller (noncallers were excluded from these analyses because by definition they did not call in the provisioning context). Because the groups were more dispersed in natural than provisioning contexts, it was more difficult in the latter context both to identify callers and to determine whether a call was a spontaneous false alarm or given in response to an eliciting stimulus. To account for the difficulty in identifying callers, we very conservatively considered calls given by unknown individuals in natural contexts to be given by every individual in the group (thus correctly counting the actual caller while conservatively increasing the call rate in natural contexts for every other individual). To account for the difficulty in determining whether calls were indeed spontaneous false alarms, we conservatively placed all calls in which we could not identify an eliciting stimulus in this category. Both Wilcoxon tests were two tailed and conducted with the VassarStats web utility (www.vassarstats.net).

The first two main predictions, i.e. those regarding the differential effects of provisioning versus nonprovisioning and the clumped versus dispersed conditions on stress hormone levels

Author's personal copy

across different alarm-calling classes, were each tested with a multilevel mixed-effects linear regression using Stata 10.0 (StataCorp LP, College Station, TX, U.S.A.). This analysis allows subjects to contribute multiple data points by including individual ID (nested in group ID in our analyses) as a random effect, and avoids the aggregation bias associated with data organized in nested structures on multiple levels (Goldstein, 2003). Details on these models are provided below.

A given fGCM sample was included in these regression analyses only if certain conditions were met. First, in the case of samples collected during the 10-day periods with provisioning experiments, the majority of the group's adults and juveniles had to have visited a feeding platform site within the 2–5 h window preceding defecation for the sample to be included in the analysis; if in such a case there was a subgroup of the main group that did not approach the platform site, we excluded the sample if the individual in question was not observed during at least one platform experiment during the stated temporal window. Second, because fGCM levels can vary by time of day (Anestis, 2010), we only included samples collected at least 1 h and 50 min after sunrise during the nonprovisioning periods, as this was the earliest we could collect samples during the periods with provisioning. Third, we excluded 50 samples from the final data set because we observed potentially stressful events (mostly intergroup encounters and possible predator encounters; see Observational Methods above) involving the individual in question in the 2–8 h window preceding defecation. This larger temporal window (relative to that used in the provisioning experiments) was used in order to be conservative regarding the potential effect of additional stressors. In addition, because any one sample may not accurately reflect GC production for a given context, we included a given individual in these analyses only if we had at least three samples from that individual that were associated with each of the two competitive conditions considered in that particular analysis and that met the criteria described above.

Owing to these considerations, two noncallers and two infrequent callers were insufficiently sampled to be included in any of the analyses, while an additional seven noncallers and two infrequent callers included in the first analysis (provisioning versus nonprovisioning conditions) were excluded from the second analysis (clumped versus dispersed conditions). For the analysis comparing periods with and without provisioning, 662 samples from 29 individuals were included in the data set, with each individual contributing a mean of 12.2 samples from periods without provisioning (range 3–29) and 10.7 samples from periods with provisioning (range 3–25). For the analysis based on food distribution within the periods with provisioning, 269 samples from 20 individuals were included in the analysis, with a mean of 6.4 samples per individual associated with the dispersed food condition (range 3–12), and 7.1 samples per individual from the clumped food condition (range 3–15).

In both of the regression models, fGCM levels were the dependent variable while the interaction between provisioning condition and the individual's alarm-calling classification (i.e. noncaller, infrequent caller or frequent caller as described above) and the main effects of these variables were the independent variables. In addition, to control for potentially confounding effects, several additional variables were included in the two models. To control for potential circadian rhythms in GC production and metabolism (Anestis, 2010), we included the time of defecation (standardized as the number of hours since sunrise; sunrise data obtained from http://aa.usno.navy.mil/) in the regression model. To control for potential effects of dominance rank (e.g. Abbott et al., 2003), we included ordinal rank as a covariate (ranks were calculated by entering all observed dyadic agonistic interactions, including aggression, spatial displacements and submissive behaviours, involving identified individuals into a dominance matrix and generating a dominance hierarchy with MatMan; Noldus Information Technology, Wageningen, The Netherlands). In addition, because GC production may differ between age and sex classes independent of other factors (Touma & Palme, 2005), we included these variables among the independent variables. To control for potential effects within and across the 10-day periods (e.g. due to habituation to the experiments), we included both the day within the 10-day period on which the sample was collected (i.e. day 1 to day 10; designated 'treatment day' in the model results) and the period within the field season (i.e. first or second half of the group's study season; designated 'period').

Finally, because fGCM values in samples collected earlier in the 2–5 h temporal window following a stressor should be higher than those collected later in this window (Wheeler, Tiddi, et al., 2013), we included among the independent variables the elapsed time since the most recent experiment falling into this 2–5 h window prior to defecation (designated 'time since experiment'). Likewise, because exposure to a greater number of stressors would be expected to increase GC values, we included the total number of experiments in which the individual participated during this 2–5 h window (designated 'number of experiments'). These two variables were included only in the analysis of the clumped versus dispersed condition; such values were, by definition, missing from samples associated with nonprovisioning conditions, and thus could not be included in the analysis comparing the provisioning and nonprovisioning conditions. Absolute time since the most recent experiment was inverse square-root transformed, while standardized time of collection of faecal samples and fGCM values were log transformed prior to analysis.

To account for possible multicollinearity among the predictor variables, we tested for correlations among all pairs of independent variables using Pearson correlations, but none were highly correlated ($r > 0.7$). We thus entered all independent variables into a full model and performed model simplification through backwards elimination (Nakagawa & Cuthill, 2007), wherein the least significant variables are sequentially deleted until all remaining variables have a P value < 0.10. Once this simplest model was achieved, the interaction of interest (i.e. experimental condition*caller type) and the main effects of these variables were re-entered into the model (if removed in the model simplification process). The significance of the interaction term and the main effects in these simplified models were used to test the first two predictions. In addition, for the purposes of plotting the results, we saved the residuals obtained from a null model that included all terms in the simplest model minus the interaction term and the main effects, and plotted these residuals against the main effects as a way to illustrate the effect of these primary variables of interest on fGCM levels beyond the effects of the control variables retained in the simplified model.

Finally, to test the third prediction, we used a two-tailed paired-samples t test using SPSS 15.0 (SPSS Inc., Chicago, IL, U.S.A.) to test whether individuals showed higher fGCM levels in samples associated with the production of spontaneous false alarms than in control samples (i.e. those associated with feeding platform experiments in which the individual did not give an alarm call). Samples associated with false alarms were limited to those collected 2–4 h following a platform experiment in which the individual had been observed to produce a spontaneous false alarm; we chose this restricted temporal window to limit samples to those that would be expected to reflect peak or near-peak fGCM values (Wheeler, Tiddi, et al., 2013). These were matched with fGCM levels in control samples collected 2–4 h following platform experiments in which the individual was known to have not given any alarm call. To control for the confounding effects of circadian rhythms, we

Author's personal copy

ensured that, within each individual, samples from each context (i.e. alarm or control) were closely matched in terms of the time of day in which they were collected. Specifically, control samples were limited to those collected within 2 h of the time of day (standardized as hours since sunrise; see above) of the time of collection of the alarm call sample, and we used the mean value of all control samples meeting this criterion. Similarly, if a given individual had multiple faecal samples associated with different alarm-calling events that were themselves closely matched in terms of their time of collection, then we used all such samples to calculate that individual's mean fGCM value for the alarm context, and also limited the controls to those that were within 2 standardized hours of all of the alarm-associated samples. Nine individuals were included in this analysis, with each individual's fGCM values being based on a mean of 1.7 alarm-associated samples (range 1–3) and 5.4 control (i.e. nonalarm) samples (range 1–10). We log transformed fGCM values prior to analysis.

RESULTS

Spontaneous False Alarm Call Rates by Context

Among the 17 individuals observed to produce spontaneous false alarm calls during feeding platform experiments, all did so at a higher rate in these provisioning contexts than in natural contexts (Wilcoxon signed-ranks test: $z = 3.61$, $N = 17$, $P < 0.001$). Within feeding platform experiments, 15 of the 17 individuals produced spontaneous false alarms more often in the clumped than the dispersed condition ($z = 3.04$, $N = 17$, $P = 0.002$). These statistically significant effects are in the same direction as in previous studies (Wheeler, 2009; Wheeler, 2010a), suggesting that the main predictions regarding the relationship between resource contestability, deceptive false alarm calling and GCs are sound for the current study period.

Provisioning versus Nonprovisioning Conditions

Overall, fGCM levels were lower in association with provisioning experiments (mean ± SD of untransformed data: 627 ± 1101 ng/g) than during periods without provisioning (1173 ± 1478 ng/g). Patterns were similar for frequent callers (provisioning: 753 ± 1409 ng/g; no provisioning: 1408 ± 1792 ng/g), infrequent callers (provisioning: 508 ± 561 ng/g; no provisioning: 829 ± 1067 ng/g) and noncallers (provisioning: 551 ± 909 ng/g; no provisioning: 1187 ± 1385 ng/g). The interaction between caller type and presence/absence of provisioning thus showed no significant association with fGCM levels in the simplified model (mixed-effects linear regression: $z = 0.89$, $N = 29$ individuals, $P = 0.374$), whereas the effect of presence or absence of provisioning was highly significant ($z = -9.67$, $N = 29$ individuals, $P < 0.001$) but in the opposite direction than predicted (Table 1, Fig. 1; see Table A1 in the Appendix for the results of the full model).

Clumped versus Dispersed Provisioning Conditions

Within periods with provisioning, fGCM levels overall were similar between the dispersed (620 ± 1098 ng/g) and clumped conditions (687 ± 1179 ng/g), and this was true for frequent callers (dispersed: 711 ± 1308 ng/g; clumped: 790 ± 1489 ng/g), infrequent callers (dispersed: 586 ± 707 ng/g; clumped: 496 ± 459 ng/g) and noncallers (dispersed: 519 ± 961 ng/g; clumped: 636 ± 890 ng/g). However, the simplified regression model controlling for the effects of sex, time of defecation and elapsed time since the most recent provisioning experiment demonstrated that fGCM levels were significantly higher in the clumped than dispersed condition ($z = 2.52$, $N = 20$ individuals, $P = 0.012$; Table 2, Fig. 2; see Table A2 in the Appendix for results of the full model). The interaction between caller type and food distribution did not significantly predict fGCM levels ($z = -1.65$, $N = 20$ individuals, $P = 0.100$), although there was a trend in the opposite direction than predicted, with the effects of the experimental manipulations on fGCM levels being strongest in noncallers and weakest in frequent callers (Table 2, Fig. 2).

Table 1
Results of the simplified multilevel mixed-effects linear regression comparing faecal glucocorticoid metabolite levels (dependent variable) during periods with and without provisioning

Variable	Coefficient	SE	*z*	*P*
Condition*	−0.431	0.045	−9.67	<0.001
Caller category	−0.023	0.041	−0.55	0.580
Interaction†	0.030	0.034	0.89	0.374
Dominance rank	−0.011	0.006	−1.70	0.090
Sex	−0.178	0.061	−2.94	0.003
Collection time (log)	−0.988	0.091	−10.86	<0.001
Period	0.071	0.030	2.41	0.016
Constant	3.783	0.159	23.82	<0.001

$N = 662$ observations from 29 subjects. Individual nested in group ID was included as a random factor.
* Provisioning versus no provisioning conditions.
† Interaction between caller category and condition.

Calling versus Noncalling Contexts

Last, among the nine individuals for whom fGCM samples associated with the production of a spontaneous alarm call were available, six showed higher fGCM levels in samples associated with false alarms than in control samples (i.e. not associated with an alarm call), while three showed the opposite effect, a

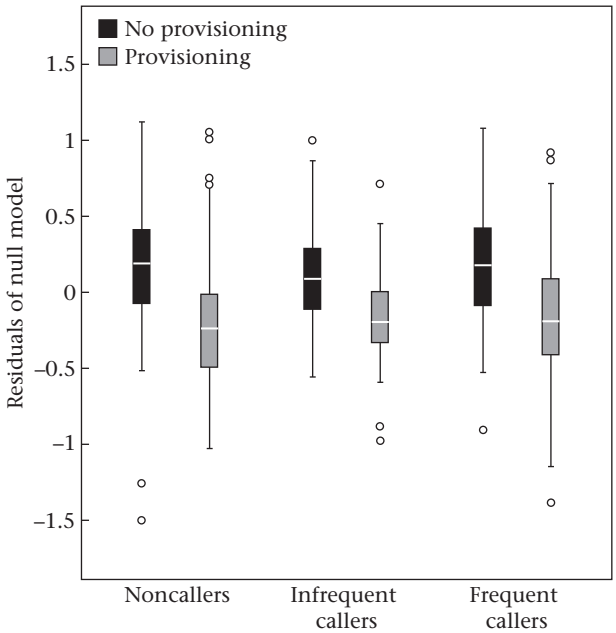


Figure 1. Box plots showing faecal glucocorticoid metabolite (fGCM) levels in tufted capuchin monkeys during periods with and without provisioning and for individuals with different propensities to produce deceptive false alarm calls. Values on the y-axis are the residuals of a null model that included only the control variables included in the reduced model in Table 1 (i.e. all variables except caller type, experimental condition and the interaction between these two variables; see Methods). Box plots show median (white line), first and third quartiles (box), range excluding outliers (whiskers), and outliers falling more than 1.5 box lengths from the edge of the box (circles). $N = 662$ fGCM samples from 29 individuals.

Author's personal copy

Table 2
Results of the simplified multilevel mixed-effects linear regression comparing faecal glucocorticoid metabolite levels (dependent variable) between the clumped and dispersed provisioning conditions

Variable	Coefficient	SE	*z*	*P*
Condition*	0.186	0.074	2.52	0.012
Caller category	−0.007	0.049	−0.14	0.890
Interaction†	−0.084	0.051	−1.65	0.100
Sex	−0.279	0.073	−3.82	<0.001
Collection time (log)	−1.475	0.155	−9.53	<0.001
Time since experiment (inverse square root)	−0.965	0.401	−2.40	0.016
Constant	4.445	0.321	13.84	<0.001

$N = 269$ observations from 20 subjects. Individual nested in group ID was included as a random factor.
* Clumped versus dispersed conditions.
† Interaction between caller category and condition.

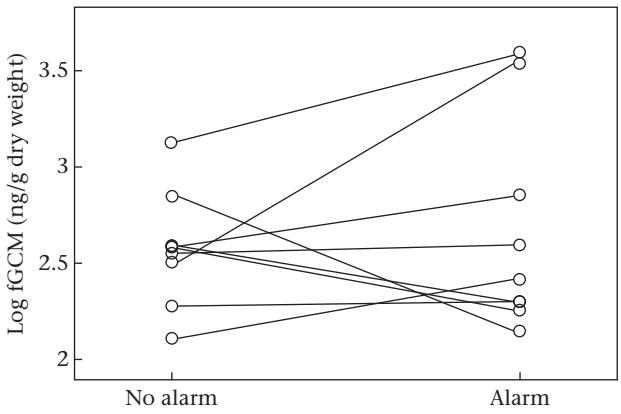


Figure 3. Line graphs showing matched comparisons of faecal glucocorticoid metabolite (fGCM) levels for nine tufted capuchin monkeys in samples associated with instances in which they spontaneously produced a false alarm call during a provisioning experiment and instances in which they did not.

nonsignificant difference (paired-samples *t* test: $t_9 = 0.539$, $P = 0.604$; Fig. 3).

DISCUSSION

None of the predictions regarding the relationship between feeding competition, adrenocortical activity and the production of deceptive false alarms were supported. First, although fGCM levels differed between periods with and without provisioning, the observed effect was in the opposite direction than predicted; fGCM levels were significantly higher during periods without provisioning, when the potential for within-group contest competition is relatively low, than during periods with controlled provisioning experiments in which foods were highly contestable. Furthermore, the effect was similar for frequent deceptive false alarm callers, infrequent callers and noncallers. Second, while analysis of fGCM levels within the provisioning contexts showed significantly higher levels in the clumped food condition than in the dispersed condition as expected, there was a nonsignificant tendency for this effect to be strongest in noncallers and weakest in frequent callers (a trend in the opposite direction than predicted). Finally, among individuals observed to spontaneously produce false alarms, fGCM levels were not higher on those occasions in which they produced such an alarm during a feeding platform experiment than during experiments in which they did not call. Taken together, these results indicate that the relationship between deceptive false alarms and the contestability of resources in tufted capuchin monkeys cannot be explained by a simple causal relationship between circulating GC levels and the production of predator-associated vocalizations (see also Cockrem & Silverin, 2002; Mazzini, Townsend, Virányi, & Range, 2013; Wilson, McDonald, & Evans, 2010).

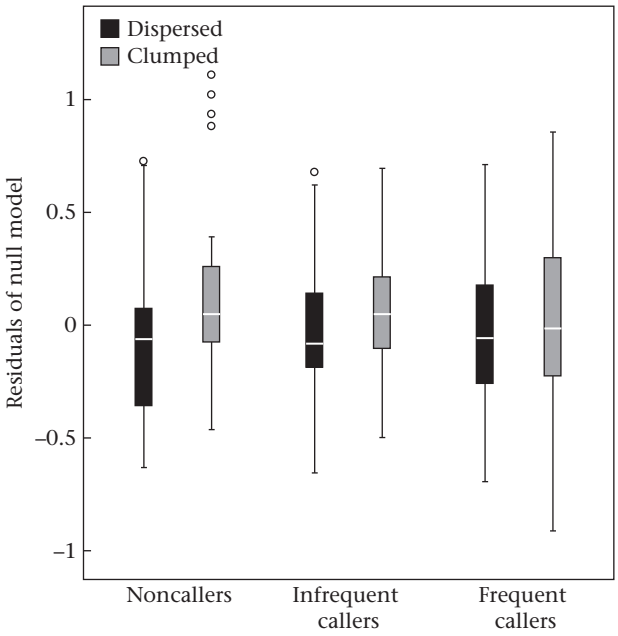


Figure 2. Box plots showing faecal glucocorticoid metabolite (fGCM) levels in tufted capuchin monkeys during provisioning periods in which food was either clumped or dispersed and for individuals with different propensities to produce deceptive false alarm calls. Values on the y-axis are the residuals of a null model that included only the control variables included in the reduced model in Table 2 (i.e. all variables except caller type, experimental condition and the interaction between these two variables; see Methods). Values greater than zero indicate fGCM levels higher than predicted by the null model, while those less than zero are lower than predicted. Box plots as in Fig. 1. $N = 269$ fGCM samples from 20 individuals.

There are seemingly two possible explanations for the observed increase in GC output during periods in which feeding platform experiments were not conducted. First, it may have been a consequence of psychosocial stress associated with the expectation of finding a high-value food (given that food was provided at feeding platforms in the preceding 10-day period), but with such expectations not being met (Ulyan et al., 2006). To test this possibility, we conducted a post hoc analysis in which we compared fGCM levels between the first 3 days of the 10-day periods without provisioning (when such expectations should be highest) and the last 3 days of such periods (when such expectations should have dissipated), but found no effect (see Appendix and Table A3). Given these results, a second possible explanation, that elevated GC levels were a response to metabolic stress associated with relative food scarcity, seems more likely. This negative relationship between food abundance and GC output has been documented in primates and other animals (Foley, Papageorge, & Wasser, 2001; Muller & Wrangham, 2004; Pride, 2005; Romero, 2002), and probably results from the fact that a major function of GCs is to increase the amount of energy available to the body in the form of glucose during times of food restriction (Reeder & Kramer, 2005; Romero, 2002).

Although deceptive false alarms do not appear to be driven simply by elevated GC levels, it is possible that metabolic stress (mediated by resource intake) and psychosocial stress (mediated by the psychological reaction to the competitive context) manifest themselves behaviourally in distinct ways. If this is indeed the case, then false alarm calls could be causally related to only psychosocial but not metabolic stress. Furthermore, high levels of metabolic

Author's personal copy

stress during periods without provisioning experiments could have masked increases in GC output resulting from psychosocial stress associated with the feeding experiments. However, this does not seem likely for two reasons. First, in such a case one would expect that frequent callers, relative to noncallers, would experience a markedly smaller decrease in fGCM levels in the provisioning condition; in contrast, the physiological responses to the experimental treatments were similar across individuals, regardless of their propensity to produce deceptive false alarms, as indicated by the lack of a significant effect of the interaction between caller type and the first provisioning condition (i.e. provisioning versus no provisioning). Second, because dominant males (who were all noncallers) have been shown to successfully monopolize platforms in the clumped condition (Di Bitetti & Janson, 2001; Janson, 1996), metabolic as well as psychosocial stress would be expected to be higher among frequent false alarm callers in the clumped relative to the dispersed condition. In contrast, although fGCM levels were indeed higher in the clumped relative to the dispersed condition, this effect tended to be weaker among frequent callers than noncallers. It is thus highly unlikely that the effect of one type of stress masked the effect of the other in this case, and it seems safe to rule out GC production resulting only from psychosocial stress as an underlying causal factor in the production of deceptive false alarms among capuchin monkeys.

Despite the fact that GCs do not seem to play a role in explaining the production of deceptive false alarm calls in capuchins, it remains highly improbable that the behaviour is performed with the intention of changing the beliefs of conspecifics regarding the presence of terrestrial predators. Such second-order intentional deception requires an ability to ascribe mental states to others (Shettleworth, 2010), an ability that is, at best, rare among nonhuman primates (Crockford, Wittig, Mundry, & Zuberbühler, 2012; Hare, Call, Agnetta, & Tomasello, 2000; but see Penn & Povinelli, 2007) and appears to be absent in at least some respects among capuchin monkeys (Hare, Addessi, Call, Tomasello, & Visalberghi, 2003). Furthermore, while it has been argued that a lack of association between vocal behaviour and GCs indeed suggests flexible call production and allows emotional mechanisms to be ruled out more generally (Mazzini et al., 2013), we would contend that we cannot in fact rule out zero-order intentionality based on such limited data. Rather, the possibility that the behaviour may be causally related to emotional mechanisms that operate independently from GC production remains open. For example, it has been argued that anxiety may be associated more with the production of catecholamines than with that of GCs (see Higham, MacLarnon, Heistermann, Ross, & Semple, 2009). It is thus possible that deceptive false alarms (or other call types) in capuchins and other animals are indeed underpinned by such an emotional mechanism, but that GC output does not provide an accurate measure of the relevant emotional state.

Although second-order intentional deception appears unlikely and simple mechanisms suggestive of zero-order intentionality cannot be ruled out completely, mechanisms characterized by first-order intentionality are possible and might be the most likely remaining explanation given the current results. For example, although the neurobiology of vocal production in primates and most other mammals suggests that the contexts in which a particular vocalization is given are largely innate and linked to specific emotional states (Hammerschmidt & Fischer, 2008; Owren et al., 1992; Seyfarth & Cheney, 2010), it remains possible that capuchins produce deceptive false alarms during feeding because they have learned that this behaviour often results in access to food. If so, then deceptive calling may involve the first-order intention of changing the action of targets. Indeed, there is limited evidence that some primates, including macaques and gibbons, can learn to spontaneously produce particular call types for the purpose of receiving food rewards (Hage, Gavrilov, & Nieder, 2013; Koda, Oyakawa, Kato, & Masataka, 2007). In the case of macaques, however, the calls that the individuals were trained to spontaneously produce are known to be food-associated (although not strictly; see Hauser & Marler, 1993), while in the case of the gibbons the extent to which the call may be normally food-associated is unknown (H. Koda, personal communication). It is thus unclear whether these cases demonstrate that nonvocal learners such as primates are indeed capable of learning to spontaneously produce a vocalization that does not have some innately grounded association with food, such as a predator- or disturbance-associated call, for the purposes of obtaining a food reward. Furthermore, it is unknown whether the ability demonstrated in macaques and gibbons to spontaneously produce specific calls is limited to those taxa with relatively sophisticated cognitive machinery, like the anthropoid primates, or is more taxonomically widespread. Testing such possibilities, in addition to potential physiological mechanisms such as those tested here, in other taxa known to use deceptive alarm calls would potentially provide insight into the extent to which apparent tactical deception across taxa is underpinned by hardwired versus more flexible mechanisms. Especially illuminating would be comparisons of the mechanisms underpinning deceptive alarm calling among vocal learners (e.g. Flower, 2011) and nonvocal learners (e.g. Bro-Jørgensen & Pangle, 2010).

Among nonvocal learners, such as nonhuman primates and most other terrestrial mammals, the widespread evidence that learning plays a more important role in determining how receivers respond to a signal than in determining the contexts in which they use those same signals (Seyfarth & Cheney, 2010; Wheeler & Fischer, 2012) suggests that counterdeception by receivers (see Gouzoules, Gouzoules, & Miller, 1996; Wheeler, 2010a; Wheeler & Hammerschmidt, 2013) probably involves more cognitive complexity than does the deceptive behaviour itself. Indeed, this may be true of tactical deception and counterdeception more generally (Byrne & Whiten, 1990), although more research is clearly needed in this regard.

In conclusion, the mechanisms underpinning tactical deception in the alarm call system of tufted capuchins remain unclear. The possibility that tactical deception in the capuchin alarm call system has primarily cognitive underpinnings remains possible, although mechanisms that require zero-order intentionality cannot yet be ruled out. An ability of callers to make learned associations between the production of terrestrial predator alarm calls and access to desired food items, suggesting a system characterized by first-order intentionality, would seem to be the most plausible explanation relying on a degree of cognitive complexity. Further experimentation will be necessary to test whether capuchin monkeys can indeed learn to voluntarily produce alarm calls for food rewards, and thus whether such associative learning in the realm of vocal usage provides a plausible cognitive path towards tactical deception in this species.

Acknowledgments

The study was funded by the US National Science Foundation (IRFP grant no. 965074 to B.C.W.). Permission to conduct the research was provided by the CIES and the Delegación Tecnica Regional NEA of the Argentine Administration of National Parks (permit no. NEA 142). We thank Andreas Koenig, Amy Lu and especially Julia Fischer for helpful advice at various stages of the study. Julia Fischer, Tabitha Price, Christof Neumann, Gabriele Schino, Christian Schloegl and two anonymous referees gave statistical advice and/or provided comments that greatly improved the manuscript. We are grateful to Andrea Heistermann for performing the hormone analyses. The study would not have been possible without the years of research in Iguazú by Charles Janson,

Author's personal copy

nor the indispensible field assistance provided by many, especially Fermino Silva, Martin Fahy, Emanuel Galetto, Elizabeth Gonzales Valentín and Anna Kordek.

References

Abbott, D., Keverne, E., Bercovitch, F., Shively, C., Mendoza, S. P., Saltzman, W., et al. (2003). Are subordinates always stressed? A comparative analysis of rank differences in cortisol levels among primates. *Hormones and Behavior, 43*, 67–82.

Anestis, S. F. (2010). Hormones and social behavior in primates. *Evolutionary Anthropology, 19*, 66–78.

Arlet, M. E., & Isbell, L. A. (2009). Variation in behavioral and hormonal responses of adult male gray-cheeked mangabeys (*Lophocebus albigena*) to crowned eagles (*Stephanoaetus coronatus*) in Kibale National Park, Uganda. *Behavioral Ecology and Sociobiology, 63*, 491–499.

Barton, R. A. (2012). Embodied cognitive evolution and the cerebellum. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*, 2097–2107.

Bercovitch, F. B., Hauser, M. D., & Jones, J. H. (1995). The endocrine stress response and alarm vocalizations in rhesus macaques. *Animal Behaviour, 49*, 1703–1706.

Blumstein, D. T., Patton, M. L., & Saltzman, W. (2006). Faecal glucocorticoid metabolites and alarm calling in free-living yellow-bellied marmots. *Biology Letters, 2*, 29–32.

Boinski, S., Gross, T. S., & Davis, J. K. (1999). Terrestrial predator alarm vocalizations are a valid monitor of stress in captive brown capuchins (*Cebus apella*). *Zoo Biology, 18*, 295–312.

Bro-Jørgensen, J., & Pangle, W. M. (2010). Male topi antelopes alarm snort deceptively to retain females for mating. *American Naturalist, 176*, E33–E39.

Brown, A., & Zunino, G. (1990). Dietary variability in *Cebus apella* in extreme habitats: evidence for adaptability. *Folia Primatologica, 54*, 187–195.

Bshary, R., Oliveira, R. F., & Grutter, A. S. (2011). Short-term variation in the level of cooperation in the cleaner wrasse *Labroides dimidiatus*: implications for the role of potential stressors. *Ethology, 117*, 246–253.

Byrne, R. W., & Corp, N. (2004). Neocortex size predicts deception rate in primates. *Proceedings of the Royal Society B: Biological Sciences*, 1693–1699.

Byrne, R. W., & Whiten, A. (1990). Tactical deception in primates: the 1990 database. *Primate Report, 27*, 1–101.

Byrne, R. W., & Whiten, A. (1988). *Machiavellian intelligence*. Oxford, U.K.: Oxford University Press.

Clinchy, M., Sheriff, M. J., & Zanette, L. Y. (2013). Predator-induced stress and the ecology of fear. *Functional Ecology, 27*, 56–65.

Cockrem, J., & Silverin, B. (2002). Sight of a predator can stimulate a corticosterone response in the great tit (*Parus major*). *General and Comparative Endocrinology, 125*, 248–255.

Crockford, C., Wittig, R. M., Mundry, R., & Zuberbühler, K. (2012). Wild chimpanzees inform ignorant group members of danger. *Current Biology, 22*, 142–146.

Cross, N., & Rogers, L. J. (2006). Mobbing vocalizations as a coping response in the common marmoset. *Hormones and Behavior, 49*, 237–245.

Di Bitetti, M. S., & Janson, C. H. (2001). Social foraging and the finder's share in capuchin monkeys, *Cebus apella*. *Animal Behaviour, 62*, 47–56.

Dunbar, R., & Shultz, S. (2007). Evolution in the social brain. *Science, 317*, 1344–1347.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178–190.

Fichtel, C., & Kappeler, P. (2002). Anti-predator behavior of group-living Malagasy primates: mixed evidence for a referential alarm call system. *Behavioral Ecology and Sociobiology, 51*, 262–275.

Flower, T. (2011). Fork-tailed drongos use deceptive mimicked alarm calls to steal food. *Proceedings of the Royal Society B: Biological Sciences, 278*, 1548–1555.

Foerster, S., & Monfort, S. L. (2010). Fecal glucocorticoids as indicators of metabolic stress in female Sykes' monkeys (*Cercopithecus mitis albogularis*). *Hormones and Behavior, 58*, 685–697.

Foley, C. A. H., Papageorge, S., & Wasser, S. K. (2001). Noninvasive stress and reproductive measures of social and ecological pressures in free-ranging African elephants. *Conservation Biology, 15*, 1134–1142.

Goldstein, H. (2003). *Multilevel statistical models*. London: Edward Arnold.

Gouzoules, H., Gouzoules, S., & Miller, K. (1996). Skeptical responding in rhesus monkeys (*Macaca mulatta*). *International Journal of Primatology, 17*, 549–568.

Hage, S. R., Gavrilov, N., & Nieder, A. (2013). Cognitive control of distinct vocalizations in rhesus monkeys. *Journal of Cognitive Neuroscience, 25*, 1692–1701.

Hammerschmidt, K., & Fischer, J. (2008). Constraints in primate vocal production. In U. Griebel, & K. Oller (Eds.), *The evolution of communicative creativity: Complexity, creativity, and adaptability in human and animal communication* (pp. 93–119). Cambridge, MA: MIT Press.

Hare, B., Addessi, E., Call, J., Tomasello, M., & Visalberghi, E. (2003). Do capuchin monkeys, *Cebus apella*, know what conspecifics do and do not see? *Animal Behaviour, 65*, 131–142.

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 59*, 771–785.

Hauser, M., & Marler, P. (1993). Food-associated calls in rhesus macaques (*Macaca mulatta*): I. Socioecological factors. *Behavioral Ecology, 4*, 194–205.

Heistermann, M., Palme, R., & Ganswindt, A. (2006). Comparison of different enzyme-immunoassays for assessment of adrenocortical activity in primates based on fecal analysis. *American Journal of Primatology, 68*, 257–273.

Higham, J. P., MacLarnon, A. M., Heistermann, M., Ross, C., & Semple, S. (2009). Rates of self-directed behaviour and faecal glucocorticoid levels are not correlated in female wild olive baboons (*Papio hamadryas anubis*). *Stress, 12*, 526–532.

Hohmann, G., Mundry, R., & Deschner, T. (2009). The relationship between sociosexual behavior and salivary cortisol in bonobos: tests of the tension regulation hypothesis. *American Journal of Primatology, 73*, 223–232.

Janson, C., Baldovino, M. C., & Bitetti, M. (2012). The group life cycle and demography of brown capuchin monkeys (*Cebus* [*apella*] *nigritus*) in Iguazú National Park, Argentina. In P. M. Kappeler, & D. P. Watts (Eds.), *Long-term field studies of primates* (pp. 185–212). Heidelberg: Springer.

Janson, C. H. (1996). Toward an experimental socioecology of primates: examples from Argentine brown capuchin monkeys (*Cebus apella nigritus*). In M. A. Norconk, A. L. Rosenberger, & P. A. Garber (Eds.), *Adaptive radiations of neotropical primates* (pp. 309–325). New York: Plenum Press.

Johnstone, R. A., & Grafen, A. (1993). Dishonesty and the handicap principle. *Animal Behaviour, 46*, 759–764.

Koda, H., Oyakawa, C., Kato, A., & Masataka, N. (2007). Experimental evidence for the volitional control of vocal production in an immature gibbon. *Behaviour, 144*, 681–692.

Koenig, A. (2002). Competition for resources and its behavioral consequences among female primates. *International Journal of Primatology, 23*, 759–783.

Mateo, J. M. (2008). Inverted-U shape relationship between cortisol and learning in ground squirrels. *Neurobiology of Learning and Memory, 89*, 582–590.

Mateo, J. M. (2010). Alarm calls elicit predator-specific physiological responses. *Biology Letters, 6*, 623–625.

Maynard Smith, J., & Harper, D. (2003). *Animal signals*. Oxford, U.K.: Oxford University Press.

Mazzini, F., Townsend, S. W., Virányi, Z., & Range, F. (2013). Wolf howling is mediated by relationship quality rather than underlying emotional stress. *Current Biology, 23*, 1677–1680.

Møller, A. (1988). False alarm calls as a means of resource usurpation in the great tit *Parus major*. *Ethology, 79*, 25–30.

Monclús, R., Rödel, H. G., Palme, R., Von Holst, D., & de Miguel, J. (2006). Non-invasive measurement of the physiological stress response of wild rabbits to the odour of a predator. *Chemoecology, 16*, 25–29.

Muller, M. N., & Wrangham, R. W. (2004). Dominance, cortisol and stress in wild chimpanzees (*Pan troglodytes schweinfurthii*). *Behavioral Ecology and Sociobiology, 55*, 332–340.

Munn, C. (1986). Birds that 'cry wolf'. *Nature, 319*, 143–145.

Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological Reviews, 82*, 591–605.

Ostner, J., Kappeler, P., & Heistermann, M. (2008). Androgen and glucocorticoid levels reflect seasonally occurring social challenges in male redfronted lemurs (*Eulemur fulvus rufus*). *Behavioral Ecology and Sociobiology, 62*, 627–638.

Owren, M. J., Dieter, J. A., Seyfarth, R. M., & Cheney, D. L. (1992). Evidence of limited modification in the vocalization of cross-fostered rhesus (*Macaca mulatta*) and Japanese (*M. fuscata*) macaques. In T. Nishida, W. McGrew, P. Marler, M. Pickford, & F. de Waal (Eds.), *Human origins* (pp. 257–270). Tokyo, Japan: University of Tokyo Press.

Palme, R., Touma, C., Arias, N., Dominchin, M. F., & Lepschy, M. (2013). Steroid extraction: get the best out of faecal samples. *Veterinary Medicine Austria, 100*, 238–246.

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B, 362*, 731–744.

Pride, R. E. (2005). Optimal group size and seasonal stress in ring-tailed lemurs (*Lemur catta*). *Behavioral Ecology, 16*, 550–560.

Reeder, D. A. M., & Kramer, K. M. (2005). Stress in free-ranging mammals: integrating physiology, ecology, and natural history. *Journal of Mammalogy, 86*, 225–235.

Romero, L. M. (2002). Seasonal changes in plasma glucocorticoid concentrations in free-living vertebrates. *General and Comparative Endocrinology, 128*, 1–24.

Schwarzenberger, F., Möstl, E., Palme, R., & Bamberg, E. (1996). Faecal steroid analysis for non-invasive monitoring of reproductive status in farm, wild and zoo animals. *Animal Reproduction Science, 42*, 515–526.

Searcy, W. A., & Nowicki, S. (2005). *The evolution of animal communication: Reliability and deception in signaling systems*. Princeton, NJ: Princeton University Press.

Seyfarth, R. M., & Cheney, D. L. (2010). Production, usage and comprehension in animal vocalizations. *Brain and Language, 115*, 92–100.

Shettleworth, S. J. (2010). *Cognition, evolution, and behavior*. Oxford University Press.

Soares, M. C., Bshary, R., Fusani, L., Goymann, W., Hau, M., Hirschenhauser, K., & et al. (2010). Hormonal mechanisms of cooperative behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*, 2737–2750.

Tamura, N. (1995). Postcopulatory mate guarding by vocalization in the Formosan squirrel. *Behavioral Ecology and Sociobiology, 36*, 377–386.

Touma, C., & Palme, R. (2005). Measuring fecal glucocorticoid metabolites in mammals and birds: the importance of validation. *Annals of the New York Academy of Sciences, 1046*, 54–74.

Ulyan, M. J., Burrows, A. E., Buzzell, C. A., Raghanti, M. A., Marcinkiewicz, J. L., & Phillips, K. A. (2006). The effects of predictable and unpredictable feeding schedules on the behavior and physiology of captive brown capuchins (*Cebus apella*). *Applied Animal Behaviour Science, 101*, 154–160.

Author's personal copy

Wheeler, B. C. (2009). Monkeys crying wolf? Tufted capuchin monkeys use antipredator calls to usurp resources from conspecifics. *Proceedings of the Royal Society B: Biological Sciences, 276*, 3013–3018.
Wheeler, B. C. (2010a). Decrease in alarm call response among tufted capuchin monkeys in competitive feeding contexts: possible evidence for counterdeception. *International Journal of Primatology, 31*, 665–675.
Wheeler, B. C. (2010b). Production and perception of situationally variable alarm calls in wild tufted capuchin monkeys (*Cebus apella nigritus*). *Behavioral Ecology and Sociobiology, 64*, 989–1000.
Wheeler, B. C., & Fischer, J. (2012). Functionally referential signals: a promising paradigm whose time has passed. *Evolutionary Anthropology, 21*, 195–205.
Wheeler, B. C., & Hammerschmidt, K. (2013). Factors underpinning receiver responses to deceptive false alarm calls in wild tufted capuchin monkeys: is it counterdeception? *American Journal of Primatology, 75*, 715–725.
Wheeler, B. C., Scarry, C. J., & Koenig, A. (2013). Rates of agonism among female primates: a cross-taxon perspective. *Behavioral Ecology, 24*, 1369–1380.
Wheeler, B. C., & Tiddi, B. (n.d.). Ad libitum and focal animal observations of agonism among tufted capuchins during provisioning and in natural contexts, 2010–2011. Unpublished raw data.
Wheeler, B. C., Tiddi, B., Kalbitzer, U., Visalberghi, E., & Heistermann, M. (2013). Methodological considerations in the analysis of fecal glucocorticoid metabolites in tufted capuchin monkeys (*Cebus apella*). *International Journal of Primatology, 34*, 879–898.
Wilson, D., McDonald, P., & Evans, C. (2010). Mechanisms of mate investment in the polygamous fowl, *Gallus gallus*. *Ethology, 116*, 755–762.

Appendix

Do False Expectations Increase Adrenocortical Activity?

Because analyses showed significantly higher faecal glucocorticoid metabolite (fGCM) levels in association with periods without provisioning with feeding platforms relative to periods with provisioning, in the opposite direction to our a priori predictions, we conducted a post hoc analysis to determine whether this could be explained as an increase in psychosocial stress resulting from the (false) expectation of being provisioned. Such expectations seem likely given that nonprovisioning periods consistently followed 10-day periods in which the monkeys were provided with banana pieces. If this indeed explains the higher GC output, we predicted that fGCM levels would be higher on the first 3 days of the nonprovisioning periods (when such expectations should be highest and the monkeys regularly visited the platform sites) than on the final 3 days of the nonprovisioning periods, when the monkeys had already had at least 7 days without provisioning and such expectations should be decreased. This was tested with a multilevel mixed-effects linear regression in which the log of fGCM levels was the dependent variable and day (i.e. beginning or end of a given 10-day period without provisioning) was the independent variable. We also included the standardized time of defecation (log transformed), the period of the field season (i.e. first or second half), and the age, sex and ordinal rank of the individual as independent variables to control for potentially confounding effects on fGCM levels (see main text). Individual ID nested in study group was included as a random effect. We included samples from a given individual only if we had at least two samples from each of the beginning and end of the 10-day periods. We performed model simplification through the backwards elimination method, and then reinserted the main variable of interest (i.e. beginning/end of the 10-day period) into the simplified model (see main text).

The mean ± SD of fGCM levels during the first 3 days of nonprovisioning periods was 1402 ± 1578 ng/g compared to 1010 ± 1104 ng/g on the final 3 days, a nonsignificant difference in the model controlling for potentially confounding effects ($z = 0.41$, $N = 146$ samples from 17 individuals, $P = 0.678$; full results of the simplified model are presented in Table A3). The higher fGCM levels found during periods without provisioning thus do not seem to be attributable to psychosocial stress associated with the unmet expectation of provisioning.

Table A1
Results of the full multilevel mixed-effects linear regression comparing faecal glucocorticoid metabolite levels (dependent variable) during periods with and without provisioning

Variable	Coefficient	SE	*z*	*P*
Condition*	−0.434	0.045	−9.72	<0.001
Caller category	−0.023	0.044	−0.52	0.606
Interaction†	0.032	0.034	0.94	0.349
Dominance rank	−0.011	0.007	−1.66	0.097
Sex	−0.183	0.076	−2.40	0.016
Age	0.004	0.100	−0.04	0.971
Collection time (log)	−0.985	0.091	−10.82	<0.001
Treatment day	−0.007	0.006	−1.23	0.220
Period	0.073	0.030	2.46	0.014
Constant	3.834	0.318	12.04	<0.001

$N = 662$ observations from 29 subjects. Individual ID nested in study group was included as a random factor.
* Provisioning versus no provisioning conditions.
† Interaction between caller category and condition.

Table A2
Results of the full multilevel mixed-effects linear regression comparing faecal glucocorticoid metabolite levels (dependent variable) between the clumped and dispersed provisioning conditions

Variable	Coefficient	SE	*z*	*P*
Condition*	0.172	0.074	2.33	0.020
Caller category	−0.058	0.060	−0.96	0.336
Interaction†	−0.072	0.051	−1.41	0.159
Dominance rank	0.001	0.009	0.12	0.904
Sex	−0.368	0.098	−3.76	<0.001
Age	−0.181	0.132	−1.38	0.168
Collection time (log)	−1.429	0.159	−9.00	<0.001
Treatment day	−0.013	0.009	−1.54	0.124
Period	0.078	0.048	1.63	0.103
No. of experiments	−0.022	0.023	−0.96	0.339
Time since experiment (inverse square root)	−0.986	0.407	−2.42	0.015
Constant	4.940	0.514	9.61	<0.001

$N = 269$ observations from 20 subjects. Individual ID nested in study group was included as a random factor.
* Clumped versus dispersed conditions.
† Interaction between caller category and condition.

Table A3
Results of the simplified multilevel mixed-effects linear regression comparing faecal glucocorticoid metabolite levels (dependent variable) between the first 3 days of periods without provisioning and the last 3 days of such periods

Variable	Coefficient	SE	*z*	*P*
Condition*	0.025	0.059	0.41	0.678
Sex	−0.021	0.110	−0.19	0.850
Period	0.114	0.059	1.92	0.054
Time of day (log)	−0.860	0.180	−4.78	<0.001
Constant	3.259	0.251	12.97	<0.001

$N = 146$ observations from 17 individuals. Individual ID nested in study group was included as a random factor.
* First versus last days of nonprovisioning periods.