# Do apes know that they could be wrong?

**Josep Call**

**Abstract** When confronted with uncertain or incomplete information in decision-making situations, monkeys and apes opt for either escaping the situation or seeking additional information. These responses have been interpreted as evidence of metacognitive abilities. However, this interpretation has been challenged. On the one hand, studies using the information-seeking paradigm have been criticized because subjects may simply engage in a search for information routine (e.g., search until spot the reward) without any metacognitive involvement. On the other hand, studies using the escape response paradigm have been criticized because subjects may not recognize their own state of uncertainty but have learned to use the escape response in the presence of certain stimuli configurations that create uncertainty. The current study attempted to address these two criticisms by presenting great apes (seven gorillas, eight chimpanzees, four bonobos, seven orangutans) with a seeking information task whose basic procedure consisted of presenting two hollow tubes, baiting one of them and letting subjects choose. Conditions varied depending on whether subjects had visual access to the baiting, the cost associated with seeking information, the time interval between baiting and choosing, the food quality and the additional information offered regarding the food's location. Although subjects showed a high retrieval accuracy when they had witnessed the baiting, they were more likely to check inside the tube before choosing when high stakes were involved (Experiment 3) or after a longer period of time had elapsed between the baiting and the retrieval of the reward (Experiment 2). In contrast, providing subjects with indirect auditory information about the food's location or increasing the cost of checking reduced checking before choosing (Experiment 1). Taken together, these findings suggest that subjects knew that they could be wrong when choosing.

**Keywords** Metacognition · Seeking information paradigm · Metamemory · Primates

J. Call (✉)
Max Planck Institute for Evolutionary Anthropology,
Deutscher Platz 6, 04103 Leipzig, Germany
e-mail: call@eva.mpg.de

Every time I fly abroad, I prepare my luggage the night before, and place my plane ticket and passport in my travel pouch right inside the front pocket of my carry-on bag. The next morning, just before I leave the house, I invariably check that my documents are still there. Why do I do this? Nobody has touched my bag and I remember perfectly well where I put them the previous night since I find them where I left them on the first try. My justification for this behavior is that I know that I could be wrong and I want to make sure that I am not. Interestingly, I do not check a second time that same morning, 30 s after my first check. In addition, I am less likely to check, although I still check sometimes, when I go by train, not by plane. Thus, the probability of my checking behavior is further modulated by the time elapsed since my last inspection and the cost of forgetting my documents—a train ticket being easier to replace than a plane ticket. This so-called passport effect (Call and Carpenter 2001) is just one manifestation of humans' ability to think about our memories—to engage in metacognition.

In recent years, there has been a considerable interest devoted to study metacognition in non-human animals (e.g., Call and Carpenter 2001; Foote and Crystal 2007; Hampton 2001; Inman and Shettleworth 1999; Shields et al. 1997; Smith et al. 1995, 2008; Suda-King 2008).

Animals confronted with information that they cannot disambiguate in decision-making situations opt for either escaping the situation or seeking additional information. Such responses have been interpreted as evidence for metacognitive abilities because they involve access to the individual's own psychological states (Smith et al. 1995, 2008). Smith and colleagues pioneered a method in which subjects had to perform some visual or auditory discrimination but were given the option to decline trials if they wished to do so. Rhesus macaques and humans rarely used the escape option for easy discriminations but its use sharply increased when the discrimination became increasingly difficult. These results were confirmed by additional studies showing that rats, dolphins, rhesus monkeys, apes and humans, unlike capuchin monkeys (Beran et al. 2009) do escape trials when faced with perceptual uncertainty (Foote and Crystal 2007; Smith et al. 1997, 2006).

Several authors have used a variation of the escape paradigm to investigate whether subjects could also gauge whether they had forgotten critical information (e.g., Hampton 2001; Inman and Shettleworth 1999; Suda-King 2008). This is important because unlike the previous tasks, it would show that individuals have access to information that was not currently perceivable. For instance, Hampton (2001) presented rhesus macaques with a delayed matching to sample paradigm in which subjects, after observing the sample and experiencing a certain time delay, could choose whether they wanted to take the trial. If subjects opted for skipping the trial, they received a low-quality reward whereas if they opted for taking the trial, they were presented with four alternatives one of which matched the sample they had experienced before the delay period. Upon selecting the correct alternative, they received a high-quality reward while they received no reward and a time-out period if they selected an incorrect alternative. In some trials, subjects could choose whether or not they wanted to take the test while in other trials, they were forced to take the test (there was no escape option available). Results showed that when subjects could choose to take the test, they performed significantly better than when they were forced to take the test. Monkeys also declined the test more often when no sample had been presented prior to the delay period.

In an additional experiment, Hampton (2001) varied the time elapsed between the presentation of the sample and the presentation of the alternatives. Both monkeys decreased their accuracy in selecting the correct alternative as a function of delay—the longer they had to wait the lower their accuracy became. Most importantly, he observed that one monkey increased the proportion of declined trials as a function of delay. Thus, percent accuracy was inversely proportional to the percent of declined

trials. Hampton (2001) suggested that monkeys knew when they had forgotten the correct alternative and that was the reason why they increased the proportion of declined trials. Recently, Fujita (2009) found that a capuchin monkey deployed escape responses in a manner consistent with a metacognitive account.

One alternative explanation to metacognition is that subjects have learned to associate particular stimuli configurations with the escape key. According to this account, subjects would not make a metacognitive judgment but they learned to use the escape key to avoid difficult trials and maximize their rewards (Smith 2009; Staddon et al. 2007). However, some studies have shown that rhesus macaques were able to generalize the escape response for new problems, thus suggesting that the escape response was not learned to respond to intermediate stimuli because it was used successfully to solve new discrimination problems (Kornell et al. 2007; Smith et al. in press; Washburn et al. 2006). Recently, Smith et al. (2006) found that rhesus macaques can use the escape key even when they do not receive immediate feedback from their responses, thus making a reinforcement account of the escape response harder to maintain. However, some authors have argued that it is still possible to explain these results using account for a reinforcement history (Crystal and Foote 2009).

Call and Carpenter (2001) tried to sidestep the problems raised against the uncertainty-escape paradigm by using a paradigm that involved seeking information rather than escaping responses. Their procedure presented a naturalistic situation that did not involve training responses during the test. In particular, apes and children were confronted with two hollow tubes resting on a platform at waist level and perpendicularly oriented toward the subject so that an individual could spy the inside of the tube by leaning down and peering throughout the tube (see Fig. 1). In visible trials, the experimenter inserted a reward inside the tube on the experimenter's side while the subject watched. In hidden trials, the experimenter also baited one of the tubes but did so behind a barrier so that the subject did not know which tube held the reward. In order to get the reward, subjects had to touch the baited tube first. Results indicated that chimpanzees, orangutans, and children sought information when they were ignorant of the location of food before they made a choice. Since then, these results have been confirmed with additional chimpanzees and orangutans and extended to gorillas, bonobos and rhesus macaques (Call 2004; Hampton et al. 2004) whereas the evidence is less clear for capuchin monkeys and dogs (Basile et al. 2009; Bräuer et al. 2004).

One possible explanation for the checking response is that it constitutes a generalized search behavior that animals display when they do not have a representation of the

**Fig. 1** Basic setup of the checking inside the tubes task

reward's location (Hampton et al. 2004; Kornell et al. 2007), either because they never saw the reward or because they forgot it. According to this generalized search hypothesis, individuals engage in a variety of exploratory responses until detecting the reward. However, this hypothesis is weakened by two pieces of data. First, subjects select the correct tube (on about 17–34% of the trials depending on the species) after having looked inside the empty tube only (Call and Carpenter 2001; Call 2005). This means that subjects did not need to spy the reward inside a tube to select the correct alternative. Second, although subjects looked more often inside the tubes in the hidden than the visible trials, they still looked inside the tube in visible trials in approximately 35% of the trials. Had they forgotten the place of the reward? This seems unlikely because we used very short delays and subjects perform near ceiling in delayed response tasks with such short delays (e.g., Barth and Call 2006). Another possibility is that this represented an instantiation of the passport effect alluded to earlier. Subjects knew where the reward was, but still looked inside the tube in visible trials to make sure that they remembered correctly. We explored this possibility further in the current study.

Unlike the generalized search hypothesis or some other innate or learned heuristic procedure, one of the key features of the metacognitive account is that it can explain a greater variety of phenomena in a variety of situations. Smith (2009) has argued that uncertainty responses are metacognitive and deployed in a complex and flexible manner often in situations without any prior training. The current study put this idea to the test by using a modified version of the seeking information paradigm. Subjects were presented with information about the location of a reward but varied the type of information provided (Experiment 1),

the cost associated with seeking information (Experiment 1), the time elapsed between baiting and retrieving the reward (Experiment 2) and the value of the reward (Experiment 3). If subjects engaged in metacognitive judgments, they should be able to incorporate these different elements in a flexible manner to make informed decisions.

Experiment 1 presented subjects with the visible and hidden condition from previous studies and another condition in which the tubes were shaken but visual inspection was not allowed. We predicted that subjects, particularly those able to use auditory information, would look less in the shaken condition than in the hidden condition. Moreover, two different tests that varied the orientation of the tubes on the platform assessed whether the cost of looking inside the tubes affected the likelihood of seeking information in the various conditions. Experiment 2 presented subjects with two types of visible trials. In blocked trials, subjects were prevented from spying the inside of the tubes by a barrier at the time of choice while in unblocked trials, subjects were able to peer inside the tube before choosing. In line with the results from Hampton (2001), we predicted that the probability of checking inside the tubes before making a selection would increase with the length of the delay between baiting and choosing. Experiment 3 presented subjects with visible and hidden trials with two types of reward (high quality and low quality). We expected that subjects would be more likely to check before choosing when a high-quality rather than a low-quality reward had been deposited inside the tube, making this analogous to the passport example presented at the beginning of the introduction.

## Experiment 1: Does providing auditory information about the reward's location reduce visual checking?

According to a generalized search hypothesis, subjects visually search their surroundings until they spy the reward and then choose the correct alternative. In this experiment, we prevented subjects from acquiring visual information and offered auditory information about the rewards' location (shaken condition). If subjects were seeking knowledge, they should be able to replace visual for auditory information thus reducing their reliance from visual search. As a comparison, we also administered the visible and hidden conditions used in previous studies. We predicted that subjects, particularly those able to use auditory information, would look less in the shaken condition than in the hidden condition. Subjects received two tests (straight tubes and oblique tubes) in each of the three conditions to assess the effect that checking effort would have on the probability of checking before choosing. In the straight tube test, subjects had simply to bend down to spy the

reward whereas in the oblique tube test, subjects had to bend down and to the side while straining their eyes to see inside the tube.

## Methods

### Subjects

Five gorillas, seven chimpanzees and four bonobos housed at the Wolfgang Köhler Research Center, Leipzig Zoo (Germany) participated in this study. There were 11 females and 5 males ranging from 6 to 30 years of age (see Table 1 for additional details). All male bonobos and all the adult chimpanzees were nursery reared whereas all other subjects were mother reared. All subjects lived in social groups of various sizes, with access to indoor and outdoor areas. All subjects had participated in a study in which they had to use the noise made by food placed inside one of two plastic cups to locate the baited one (Call 2004). Three gorillas, two chimpanzees and two bonobos were able to solve this task. Subjects were individually tested in their indoor cages and were not food or water deprived.

### Materials

Two gray opaque tubes (25 cm long × 4 cm in diameter) were placed on a platform about 40 cm apart (see Fig. 1). Two blue plastic covers (15 cm × 10.5 cm) placed in front of the tubes served to block visual access to the tubes contents. Monkey biscuits were used as rewards. The test materials were presented on a sliding platform situated behind a plexiglass partition that separated the subject from the experimenter (E). This plexiglass partition had two circular holes cut on its bottom (on each side) that allowed subjects select one of the tubes by touching it.

### Procedure and design

The basic procedure was as follows. The E sat facing the subject behind the platform separated by a plexiglass partition. Two tubes rested on the platform perpendicularly oriented with respect to the partition so that subjects were able to peer inside the tubes by lowering their head to the opening of the tubes. However, visual access to the inside of the tubes was prevented by the plastic covers placed in front of the tube

**Table 1** Subjects included in the study

| Name | Species | Age (years) | Sex | Rearing history | Experiment participation |
|------|---------|-------------|-----|-----------------|--------------------------|
| Dorien | Chimpanzee | 23 | F | Nursery | 1,2,3 |
| Fraukje | Chimpanzee | 27 | F | Nursery | 1,2,3 |
| Jahaga | Chimpanzee | 10 | F | Mother | 1,2,3 |
| Fifi | Chimpanzee | 10 | F | Mother | 1,2,3 |
| Sandra | Chimpanzee | 10 | F | Mother | 1,2,3 |
| Gertruida | Chimpanzee | 10 | F | Mother | 1,2 |
| Frodo | Chimpanzee | 10 | M | Mother | 1,2,3 |
| Patrick | Chimpanzee | 6 | M | Mother | 2 |
| Bebe | Gorilla | 26 | F | Unknown | 1,2,3 |
| Ndiki | Gorilla | 26 | F | Unknown | 1,2 |
| Vimoto | Gorilla | 8 | M | Mother | 3 |
| Viringika | Gorilla | 8 | F | Mother | 1,2,3 |
| Vizuri | Gorilla | 8 | F | Mother | 3 |
| Nkwango | Gorilla | 7 | M | Mother | 1,2,3 |
| Ruby | Gorilla | 6 | F | Mother | 1,2,3 |
| Joey | Bonobo | 21 | M | Nursery | 1,2,3 |
| Ulindi | Bonobo | 10 | F | Mother | 1,2,3 |
| Limbuko | Bonobo | 8 | M | Nursery | 1,2,3 |
| Kuno | Bonobo | 7 | M | Nursery | 1,2,3 |
| Dunja | Orangutan | 30 | F | Unknown | 2,3 |
| Bimbo | Orangutan | 23 | M | Mother | 2 |
| Pini | Orangutan | 15 | F | Mother | 2,3 |
| Walter | Orangutan | 14 | M | Mother | 2,3 |
| Dokana | Orangutan | 14 | F | Mother | 2 |
| Toba | Orangutan | 9 | F | Mother | 2,3 |
| Padana | Orangutan | 6 | F | Mother | 2,3 |

openings on the subject's side. The experimenter placed a reward inside one of the tubes. In order to get the reward, subjects had to touch the baited tube. There were three conditions depending on the baiting procedure:

*Visible*: E showed a reward to the subject and placed it inside of one of the tubes on the E's side while the subject was watching.

*Hidden*: Prior to inserting the reward inside a tube, E placed the opaque barrier between the tubes and the plexiglass partition thus blocking the subject's visual access to the baiting event. Then, E showed a reward to the subject and inserted it inside of one of the tubes. The experimenter placed his fingers (and the reward) inside both tubes in succession and left the reward in one of the tubes. Then, the experimenter removed the opaque barrier from the platform.

*Shaken*: This condition was identical to the hidden condition except that after the E had removed the opaque barrier, he took, shook and replaced each tube to their original position in succession. Special care was taken to block the tubes' openings with his hands to prevent spilling the rewards or subject's visual access to the tubes' interior. When shaken, the reward placed inside the baited tube produced a clearly audible rattling sound.

Once the baiting procedure was completed (including the shaking in the shaken condition), the experimenter removed the lids blocking the tubes and pushed the platform forward to allow the subject to select one of the tubes by touching it. We conducted two tests in succession (each composed by the three conditions mentioned above) that differed in the orientation of the tubes. In the straight tubes test, tubes were parallel to each other and perpendicular to the plexiglass partition. In the oblique tube test condition, tubes formed a 60° angle with each other and with the plexiglass partition. This meant that the reward was easier to see in the straight than the oblique test. Each subject received three 12-trial sessions per test. Each session consisted of 4 trials per condition. Overall, each subject received 72 trials (2 tests × 3 sessions × 12 trials). The order of trial presentation and the location of the reward were counterbalanced every sic trials with the only two restrictions that it appeared the same number of times on each side and could not appear more than two times in succession on the same side.

*Data scoring and analysis*

We videotaped all trials and scored the tube selected and whether subjects looked inside the tube before choosing. We scored a tube selection as the first tube (or plastic lids if present) touched by the subject. We scored a response as looking inside the tubes when subjects lowered their head and body so that their eyes were aligned with the tube opening. Inter-observer reliability based on 16% of the trials was excellent in both cases (choosing: kappa = 1.0; looking: kappa = 0.87).

We classified subjects into two groups [pass ($n = 7$) vs. fail ($n = 9$)] depending on whether they had performed above chance in the auditory condition of the Call (2004, Experiment 1) study. We analyzed the percent of trials with looking responses as a function of test, condition and group. We used one-tailed non-parametric statistics (Friedman, Mann–Whitney, Wilcoxon tests) throughout due to the directional nature of our predictions.

Results

Figure 2a presents the percentage of trials in the straight tube test in which subjects looked inside the tube as a function of condition and group. Overall, there were significant differences between conditions in the straight tubes test (Friedman test: $\chi^2 = 22.12$, $df = 2$, $P < 0.001$, $N = 16$). Apes looked significantly less often in the visible condition compared to the shaken (Wilcoxon exact test: $T = 105$, $P < 0.001$, $N = 14$) and hidden conditions (Wilcoxon exact test: $T = 105$, $P < 0.001$, $N = 13$). In
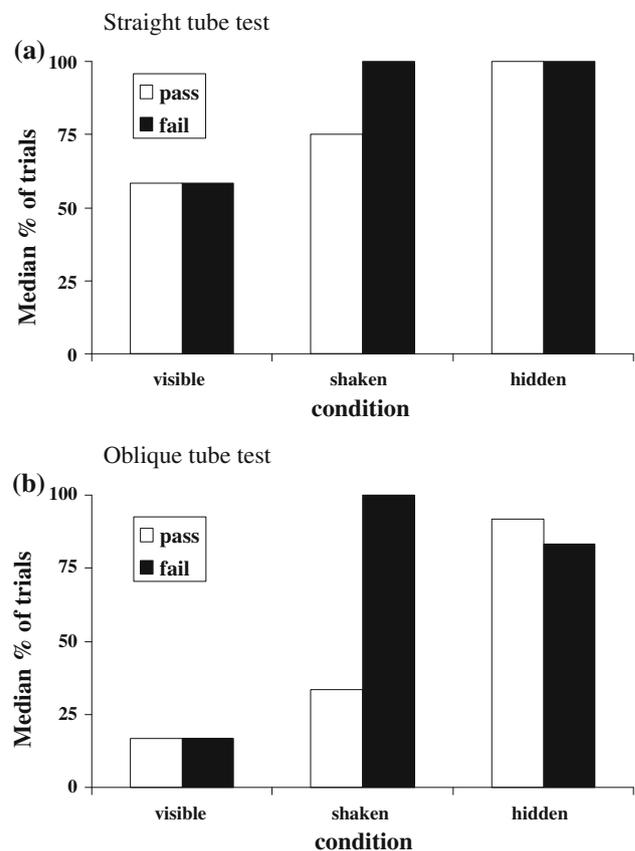


**Fig. 2** Median percentage of trials in which subjects looked inside the tube before choosing in the **a** straight and **b** oblique tube conditions as a function of group

contrast, there were no significant differences between the shaken and the hidden conditions (Wilcoxon exact test: $T = 33$, $P = 0.1016$ one-tailed, $N = 9$). There were no significant differences between groups in any of the three conditions (Mann–Whitney tests: visible: $z = 0.06$, $P = 0.96$; shaken: $z = 1.17$, $P = 0.24$; hidden: $z = 0.23$, $P = 0.82$; $N = 16$ in all cases).

Table 2 presents the median percentage of trials in which subjects looked inside the tube as a function of species and condition. There were significant differences between species in all conditions (see Table 2). However, post-hoc pairwise Mann–Whitney tests only confirmed that gorillas looked less than chimpanzees in the shaken condition ($Z = 2.87$; $P = 0.005$) and than bonobos in the visible condition ($Z = 2.35$, $P = 0.032$).

Focusing on the oblique tubes test (Fig. 2b), there were overall significant differences between conditions (Friedman test: $\chi^2 = 20.04$, $df = 2$, $P < 0.001$, $N = 16$). Apes looked significantly less often in the visible condition compared to the shaken (Wilcoxon exact test: $T = 78$, $P < 0.001$, $N = 12$) and hidden conditions (Wilcoxon exact test: $T = 91$, $P < 0.001$, $N = 13$). In contrast, there were no significant differences between the shaken and the hidden conditions (Wilcoxon exact test: $T = 29$, $P = 0.248$ one-tailed, $N = 9$). However, there were significant differences between groups in the shaken condition (Mann–Whitney test: $z = 2.03$, $P = 0.042$, $N = 16$) but not in the visible (Mann–Whitney test: $z = 0.16$, $P = 0.87$, $N = 16$) or hidden conditions (Mann–Whitney test: $z = 0.0$, $P = 1.0$, $N = 16$). There were significant differences between species in the hidden condition (see Table 2). Post-hoc pairwise Mann–Whitney tests indicated that gorillas looked less than bonobos in the hidden condition ($Z = 2.25$, $P = 0.032$).

**Table 2** Median percentage of trials in which subjects looked inside the tube before choosing in the (a) straight and (b) oblique tube conditions as a function of species

| Species | Condition | | |
|---|---|---|---|
| | Visible | Shaken | Hidden |
| Straight test | | | |
| Gorilla | 0 | 8.3 | 16.7 |
| Chimpanzee | 58.3 | 100 | 100 |
| Bonobo | 62.5 | 91.7 | 100 |
| $\chi^2_2$ ($P$) | 6.30 (0.043) | 9.65 (0.008) | 6.96 (0.031) |
| Oblique test | | | |
| Gorilla | 0 | 8.3 | 8.3 |
| Chimpanzee | 25 | 91.7 | 91.7 |
| Bonobo | 25 | 66.7 | 100 |
| $\chi^2_2$ ($P$) | 2.10 (0.35) | 2.53 (0.28) | 6.62 (0.036) |

$\chi^2_2$ ($P$) presents the values of the Kruskal–Wallis test and its associated $P$-value

Comparing the shaken and the hidden conditions within each of these two groups after pooling together the data from the straight and oblique tests confirmed this result. Subjects that passed the pretest looked inside the tube less often in the shaken than in the hidden test (Wilcoxon exact test: $T = 26$, $P = 0.0234$ one-tailed, $N = 7$). In contrast, subjects that failed the pretest looked inside the tube equally often in both conditions (Wilcoxon exact test: $T = 10$, $P = 0.3125$ one-tailed, $N = 5$).

Finally, we compared the percentage of looking responses in the straight and oblique tests for each of the conditions (compare Fig. 2a, b for each condition). Subjects looked less often in the oblique than the straight test both in the visible (Wilcoxon exact test: $T = 74$, $P = 0.006$, $N = 12$) and in the hidden conditions (Wilcoxon exact test: $T = 33.5$, $P = 0.027$, $N = 8$) but not in the shaken condition (Wilcoxon exact test: $T = 57.50$, $P = 0.14$, $N = 12$).

## Discussion

Subjects were less likely to visually inspect the tubes when they had received auditory information about the tube's contents, which means that they were able to incorporate auditory information emanating from the tube to guide their choices. However, only those subjects that had been able to use such information in a previous study were able to incorporate this information here. Moreover, subjects were less likely to look when the cost of looking was higher; something that Hampton et al. (2004) had also observed in rhesus macaques. Interestingly, in the current study the reduction of looking after increasing its cost was particularly marked when subjects had witnessed the baiting (41 vs. 12% reduction in the visible and hidden trials, respectively). This suggests that subjects knew when looking could be dispensed without a substantial loss. These results suggest that subjects did not use a fixed search sequence based on spying the food followed by choosing but could integrate auditory information and took into account the cost of looking in making informed decisions.

Although gorillas looked inside the tubes less often than chimpanzees and bonobos in some conditions, the pattern was not consistent between the straight and the oblique test. In fact, the pattern of gorillas across conditions was similar to that of chimpanzees and bonobos (i.e., the lowest level of looking in the visible condition and the highest level in the hidden condition, see Table 2).

## Experiment 2: Does forgetting predict checking?

In this experiment, we returned to the issue of seeking visual information but in this case we manipulated the

interval between baiting the tubes and selecting one of them in an attempt to foster forgetting and investigate whether looks would also increase accordingly. We borrowed Hampton's ([2001](#)) procedure and adapted it to the seeking information paradigm by introducing a variable time delay between baiting the tubes and letting subjects select one of them. Thus, we showed subjects the location where food was hidden (out of two possible locations) and then introduced a variable time delay (5, 20, 60, or 120 s) between baiting one of the tubes in full view of the subject and allowing the subject to select one of the tubes. In half of the trials (unblocked), subjects were allowed to look inside the tube while in the other half (blocked) they were not. The blocked trials allowed us to estimate the apes' forgetting curve whereas the unblocked trials allowed us to assess the probability of looking as a function of delay. We predicted that both forgetting and looking would increase as a function of the delay between baiting and choosing.

## Methods

### Subjects

Seven orangutans, eight chimpanzees, five gorillas and four bonobos housed at the Wolfgang Köhler Research Center, Leipzig Zoo (Germany) participated in this study. There were 15 females and 9 males ranging from 6 to 30 years of age (see Table 1 for additional details). All male bonobos and all the adult chimpanzees were nursery reared whereas all other subjects were mother reared. All subjects lived in social groups of various sizes, with access to indoor and outdoor areas. Subjects were individually tested in their indoor cages and were not food or water deprived.

### Materials

Two white opaque square-shaped tubes (5 cm × 5 cm × 30 cm) were placed on a platform about 40 cm apart. Two small wooden blocks (6 cm × 6 cm) placed in front of the tubes served to block visual access to the tubes contents. Grapes, pieces of banana or monkey biscuits were used as rewards. The test materials were presented on a sliding platform situated behind a plexiglass partition that separated the subject from the experimenter (E). This plexiglass partition had two circular holes cut on its bottom (on each side) that allowed subjects select one of the tubes by touching it.

### Procedure and design

The E sat facing the subject behind the platform separated by a plexiglass partition and placed the tubes on the platform perpendicularly oriented with respect to the partition. This meant that subjects were able to peer inside the tubes by lowering their head to the opening of the tubes. E then blocked the view of these openings by placing each of the two wooden blocks in front of the tube opening on the subject's side. E showed a reward to the subject and then placed it inside of one of the tubes on the E's side and waited for 5, 20, 60 or 120 s depending on the trials before pushing the platform forward and giving the subject a choice between the two tubes. There were two conditions depending on whether the experimenter removed the wooden blocks from the tube openings before pushing the platform forward to give the subject that opportunity to select one tube.

*Blocked*: E touched each of the two wooden blocks simultaneously and then simply pushed the platform forward.
*Unblocked*: E removed the wooden blocks from the tube openings before pushing the platform.

Thus, there were two conditions each with four different time delays between the baiting and selection. Each subject received three 16-trial sessions. Each session contained 8 different types of trials (presented twice during a session) resulting from the combination of two forms of presentation (open vs. closed) with four time delays (5, 20, 60 and 120 s). Overall, each subject received 48 trials (8 trial combinations × 2 repetitions × 3 sessions). The order of trial presentation and the location of the reward were counterbalanced every eight trials with the only two restrictions that it appeared the same number of times on each side and could not appear more than two times in succession on the same side.

### Data scoring and analysis

We videotaped all trials and scored them in the same way as in Experiment 1. Inter-observer reliability based on 35% of the trials was excellent in both cases (choosing: kappa = 1.0; looking: kappa = 0.80). We analyzed the percent of correct selections and the percent of trials with looking responses as a function of the delay between baiting and choosing. We used blocked trials with different delays to investigate whether subject's performance would decrease with longer delays. We used unblocked trials to examine whether subjects would be more likely to look inside the tubes with increasing delays. Finally, we calculated the correlation between percent of correct trials and percent of trials with looking responses as a function of delay. We used one-tailed non-parametric statistics (Friedman, Wilcoxon, and Spearman tests) throughout due to the directional nature of our predictions.

## Results

Figure 3 presents the percentage of correct trials in the closed condition (left $Y$-axis) as a function of delay. Subjects' performance in the closed trials significantly decreased as a function of delay (Friedman test: $\chi^2 = 6.37$, $df = 3$, $P < 0.05$) and there were no significant differences between species for any of the delays (Kruskal–Wallis tests: $\chi^2 < 5.26$, $df = 3$, $P > 0.15$ in all cases). Moreover, subjects performed significantly worse in closed trials compared to open trials (Wilcoxon test: $z = 3.38$, $P = 0.001$, $N = 22$).

Figure 3 also presents the percentage of trials in which subjects looked inside the tube before choosing in open trials (right $Y$-axis) as a function of delay. Subjects looked significantly more often the longer the delay (Friedman test: $\chi^2 = 21.2$, $df = 3$, $P < 0.001$) and again, there were no significant differences between species for any of the delays (Kruskal–Wallis tests: $\chi^2 < 3.86$, $df = 3$, $P > 0.27$ in all cases). In addition, their accuracy in spying the reward on their first attempt significantly decreased as a function of the delay, further reinforcing the idea that they may have forgotten the exact position of the reward after longer delays (Friedman test $\chi^2 = 6.59$, $df = 3$, $P < 0.05$). Finally, there was a strong negative relation between the percentage of correct trials and the percentage of trials in which subjects looked inside the tube ($r_3 = -0.99$, $P = 0.011$).

## Discussion

Increasing the time interval between observing the baiting of the containers and choosing one of them produced the two predicted effects. First, it reduced the subjects' accuracy in finding the reward (blocked trials)—an effect commonly found in the literature on memory. Second, it increased subjects' likelihood of looking inside the tubes before choosing (unblocked trials). This result is analogous
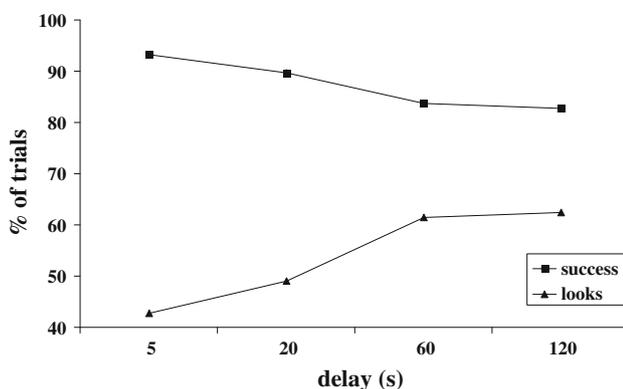


**Fig. 3** Mean percentage of trials with success (in the blocked condition) and looks (in the unblocked condition) as a function of the time elapsed between baiting and choosing

to that described in rhesus macaques in the escape response paradigm and consistent with the idea that apes knew that they had forgotten.

However, subjects also directed their first look preferentially to the baited tube including in those conditions with the longest delay. If forgetting and knowing that one has forgotten were the sole explanation for our results, subjects should not have directed their first look preferentially to the baited tube. The fact that they did so suggests that they still remembered, at least in some trials, the location of the reward. It is true that the accuracy in their first looks (just like their choice accuracy in blocked trials) showed that they remembered less well after the longer delay but it was still well above chance. This means that forgetting played a role, but this alone cannot be the sole explanation for the looking behavior. So how can this be resolved? One possibility is that some looks were motivated by a lack of information but others resulted from their attempts to make sure that they would pick the correct alternative. We tested this possibility in the next experiment.

## Experiment 3: Does checking depend on food quality?

We presented subjects the visible and hidden conditions of previous experiments but varied the quality of the reward. In some trials, we baited the tube with a high-quality reward whereas in other trials, we used a low-quality reward. The prediction is that subjects would check more often when a high-quality as opposed to a low-quality reward is at stake. Additionally, we predicted that higher levels of checking for high-quality rewards should be independent of whether subjects had or had not witnessed the baiting. We also tested whether subjects remembered the location of high-quality rewards better than the location of low-quality rewards.

### Methods

#### Subjects

Five orangutans, six chimpanzees, six gorillas and four bonobos housed at the Wolfgang Köhler Research Center, Leipzig Zoo (Germany) participated in this study. There were 14 females and 7 males ranging from 6 to 30 years of age (see Table 1 for additional details).

#### Materials

The same square-shaped tubes used in Experiment 2, two small opaque plastic cups (6 cm in diameter × 5.5 cm in height) presented on a sliding platform previously described in Experiment 2. We also used an opaque screen to

occlude the baiting in the hidden condition. Grapes and pieces of orange (or carrot) served as the high- and the low-quality rewards, respectively.

### Procedure and design

The basic procedure was the same as in previous experiments: the E sat facing the subject behind the platform separated by a plexiglass partition and presented the test materials and scored the subjects' choice. We administered a test and a control test. There were three types of trials during the test: preference, visible and hidden baiting. In the preference trials, the E simultaneously placed one high- and one low-quality reward forming a straight line and separated by 40 cm so that the rewards were aligned with the holes in the partition but still outside of the subject's reach. Once subjects had looked at the rewards, the E pushed the platform for subjects to make a choice. In visible and hidden trials, the E placed the tubes on the platform perpendicularly oriented with respect to the partition and baited them using the same method described in Experiment 1. In half of the trials, the reward was a high-quality reward whereas in the other half, it was a low-quality reward. Each subject received four 16-trial sessions. The first four trials of each session were devoted to the preference trials whereas the remaining 12 trials corresponded to visible and hidden trials with order of presentation counterbalanced across trials. Overall, each subject received 16 preference trials, 24 visible trials and 24 hidden trials. Upon completing this test, subjects received a control test in which the E placed the two opaque cups on the platform forming a straight line and separated by 40 cm. In full view of the subject, he placed a reward under one of the cups and allowed the subject to make a choice. Again, the high-quality reward was used in half of the trials and the low-quality reward in the other half. Each subject received two 12-trial sessions. The location of the reward (or the high-quality reward in case of preference tests) was counterbalanced across trials so that it appeared the same number of times on the left and the right sides.

### Data scoring and analysis

We videotaped all visible, hidden and control trials. We scored and analyzed the data in the same way as in Experiment 1. We did not assess inter-observer reliability due to the high reliability obtained in the last two experiments after using the same measures.

### Results

Subjects selected the high-quality reward significantly more often than the low-quality reward in the preference trials (Wilcoxon test: $z = 3.50$, $P < 0.001$, $N = 19$; median high-quality $= 87.5\%$). Figure 4 presents the percentage of trials in which subjects looked inside of the tube as a function of the baiting visibility (visible vs. hidden) and the reward quality (high vs. low). Overall, subjects were more likely to look inside the tube in the hidden than the visible condition (Wilcoxon test: $z = 4.02$, $P < 0.001$, $N = 21$). Moreover, apes looked significantly more often inside the tube when a high quality-rather than a low-quality reward had been deposited inside the tube (Wilcoxon test: $z = 3.06$, $P < 0.001$, $N = 19$). This difference was apparent both in visible (Wilcoxon test: $z = 2.60$, $P = 0.009$, $N = 15$) and in hidden trials (Wilcoxon exact test: $T = 93$, $P = 0.009$, $N = 14$). There were no significant differences between species in any of the four conditions (Kruskal–Wallis tests: visible-high: $\chi^2 = 3.07$, $df = 3$, $P = 0.38$; visible-low: $\chi^2 = 4.40$, $df = 3$, $P = 0.22$; hidden-high: $\chi^2 = 4.19$, $df = 3$, $P = 0.24$; hidden-low: $\chi^2 = 6.45$, $df = 3$, $P = 0.092$).

Subjects' accuracy in spying the reward on their first attempt was comparable for high- (median $= 100\%$ of the trials) and low-(median $= 100\%$ of the trials) quality rewards (Wilcoxon test: $T = 65.5/25.5$, $z = 1.40$, $P = 0.16$). In contrast, and as expected, subjects showed a much lower accuracy in hidden (median $= 41.6\%$) compared to visible (median $= 95.5\%$) trials (Wilcoxon test: $z = 4.14$, $P < 0.001$, $N = 24$). Finally, subjects were equally accurate in retrieving both types of the rewards in the control test (Wilcoxon test: $N = 1$; $N = 19$ ties; median number of trials correct: high quality $= 100\%$; low quality $= 100\%$).

### Discussion

Subjects looked more often inside a tube when a high- rather than a low-quality reward was at stake, thus confirming our prediction. Such a difference is particularly remarkable given that subjects remembered the location of
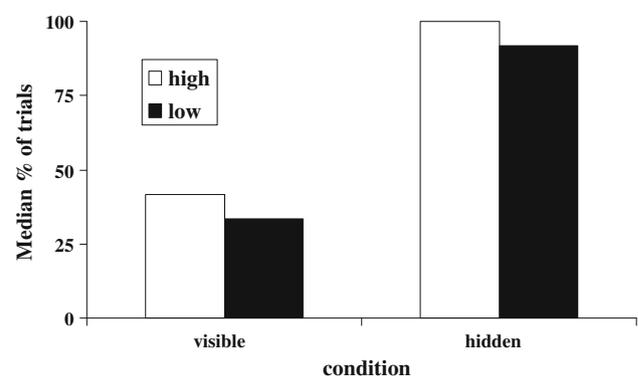


**Fig. 4** Median percentage of trials in which subjects looked inside the tube before choosing as a function of condition and food quality

the reward equally well regardless of its quality as shown by the fact that subjects directed their first look preferentially to the baited tube and the data of the control test. These latter results cast serious doubts on the counterintuitive notion that the reason subjects looked more often to the high-quality than the low-quality foods was because subjects suffered a greater memory loss for the high- compared to the low-quality food. Moreover, as expected, they did not direct their first look to the baited location in the hidden trials but they did so in the visible trials, thus replicating the results of the previous experiments.

One could argue that subjects simply preferred to look at the high-quality food because looking at it was more 'rewarding' than looking at low-quality food. However, this explanation is too narrow because it would not explain why subjects looked at high-quality rewards inside the tube more often in the hidden than the visible condition. In both cases, subjects knew that high-quality food was being placed inside the tube but they looked more often when they had not seen where the food was placed. If seeing the high-quality food inside the tube was intrinsically rewarding, they should have looked equally often in both conditions. They did not because they were seeking information about the food location to be able to choose accurately. Additionally, looking as reward cannot explain why there were more looks after longer delays compared to shorter ones and less looks after additional information had been provided.

## General discussion

Subjects were more likely to look inside the tube before choosing when the cost of looking was lower (Experiment 1), high stakes were involved (Experiment 3) and a longer period of time had elapsed between the baiting and the retrieval of the reward (Experiment 2). On the other hand, providing subjects with indirect (non-visual) information about the rewards' location reduced their looks inside the tube before they chose (Experiment 1).

It has been suggested that subjects in metacognitive experiments involving an escape response react to uncertainty without being aware of their knowledge states or their level of uncertainty (Carruthers 2009; Staddon et al. 2007). Another non-metacognitive possibility is that subjects simply react to the anxiety that they may feel in uncertain situations but without representing their state of anxiety. It is even conceivable that subjects do not react to uncertainty or anxiety per se but to certain stimuli configurations associated with certain reinforcement contingencies. In principle, the same reasoning can be applied to seeking additional information, i.e., subjects look inside the tubes more often when they are uncertain to maximize their

chances of getting the reward. Let's briefly analyze each of these explanations in light of the findings presented in the current study.

Several recent studies have challenged the idea that certain stimuli configurations associated with particular reinforcement contingencies determine the use of uncertain responses (see Smith 2009 for a review). For instance, macaques can use escape responses appropriately when presented with novel stimuli and in the absence of trial-by-trial reinforcement. The positive findings with a set of novel situations in the absence of training in the current study fit well with those of previous results. Moreover, the effect that the cost of checking or the value of the reward had on the probability of checking before choosing is difficult to explain purely in terms of stimuli configurations (and the information that is encoded from them) without proposing additional explanations.

The current data also challenge the response competition hypothesis (Hampton et al. 2004). According to this hypothesis, when the tendency to reach for the reward is strong (e.g., when subjects had seen the food location), it overrides the looking response but when the reaching tendency is weaker (e.g., when subjects have not seen the reward location), looking responses are more likely to be expressed. The current results on reward quality challenge this hypothesis because reaching for the high-quality reward, which is presumably stronger than reaching for a low-quality reward, was accompanied by an increase in looking, not a reduction as the response competition hypothesis predicts.

Non-metacognitive accounts based on postulating anxiety as the mediating variable are more successful in explaining our current results. One could argue that subjects are more likely to look inside the tube before choosing after longer delays because the representation about the food location tends to degrade over time. Moreover, subjects might be more anxious when food of a higher quality is at stake, which again would explain the increase in looking behavior as it would reduce anxiety. However, the anxiety-based account does not explain the effect that the cost of looking has on the looking responses. The data showed that the more costly the looking response, the less subjects looked, although presumably the anxiety level would predict the same level of looking in both cases. Presumably, this is so because subjects can decide based on the state of the information that they possess whether looking is needed or not. However, note that even an account based on anxiety also requires subjects to analyze the state of the information that they possess, which might require some sort of metacognitive assessment. Moreover, the anxiety-based explanation presupposes that the subject is anxious at the time of choice, but it is unclear that this is the case. It is conceivable that subjects are not currently

anxious but they might anticipate being anxious about not securing the reward in the future. This second possibility may mean that subjects are indeed contrasting their current actual states with their hypothetical future ones.

At the very least, the current results indicate that the looking response appears to be a function of at least three factors: the cost of looking inside the tube, the value of the reward and the state of the information. The combination of these three factors creates an information processing system that possesses complexity, flexibility and control, three of the features of metacognition as recently argued by Smith (2009). The current study confirms that subjects seek information more often when they have not witnessed the location of the reward or after longer delays between baiting and retrieval. More interestingly, they seek information about the food location differentially even when their uncertainty is minimal depending on the quality of the reward, the cost associated with looking and whether additional non-visual information about the reward location has been provided. It seems justified to attribute minimal uncertainty to subjects because they displayed high retrieval accuracy even without acquiring additional information and opted for not seeking unnecessary information when the cost of seeking was increased.

In previous studies, escape responses have been invariably associated with the presentation of difficult trials by virtue of the ambiguity of the stimuli presented (discrimination tasks) or the degradation of the information over time (memory tasks). Both preparations create high uncertainty. Although metacognitive judgments are often recruited precisely in situations of high uncertainty, the current study suggests that this does not need to be the case. Even situations with minimal uncertainty may be sufficient to reveal metacognitive judgments when other factors (e.g., value of the reward) come into play. One consequence of these findings is that uncertainty monitoring may be only part of the information that goes into making metacognitive judgments. One way to broadly characterize the results of the current study is that subjects are not merely uncertain, they may also know that they can be wrong (an even are capable of anxiety anticipation) particularly after a long delay between baiting and retrieval and when it matters the most (i.e., high-quality food is at stake). This would explain why looks increased in those conditions. Moreover, it would also explain why subjects can forego looking when they have already seen the food location and the cost of looking is increased.

Providing additional information about the food location (Experiment 1) would act in the opposite way by reducing the likelihood that subjects may be wrong and consequently reducing looking behavior. The fact that subjects were able to incorporate information from a different modality than the one that they would acquire through visual checking and they can engage in some inferential reasoning by exclusion in this context (Call and Carpenter 2001; Call 2005) suggests that this is a flexible system that can integrate information from different modalities. Certainly not a system that is based on following a search strategy that is solely controlled by spying the reward. Moreover, note that checking can increase for different reasons including, high stakes, elapsed time, or not witnessing the baiting. But, the principle is always the same, trying to reduce the probability of being wrong, not solely uncertainty because subjects are very accurate in their searches.

One final word about species differences. Experiment 1 produced some indication that gorillas were less likely to look inside the tubes than chimpanzees and bonobos, but this result could not be confirmed in the next two experiments. Moreover, even if gorillas checked the tubes less often than the other species in Experiment 1, they still showed the same pattern across conditions. That is, they checked less often in those conditions in which they had witnessed the baiting of the tube. Yet, it is tantalizing to think that gorillas may possess weaker metacognitive abilities compared to chimpanzees and bonobos, particularly when considering the data suggesting that gorillas are less likely to pass the mirror self-recognition test (e.g., Suarez and Gallup 1981, but see Posada and Colell 2007). Based on the current evidence, however, such conclusion would be premature both because the current study only offers weak support for it and because previous studies (Call 2005) had failed to detect a difference between gorillas and the other great apes. Nevertheless, given that some primates such as capuchin monkeys might be less metacognitive than macaques (Beran et al. 2009; Smith 2009; but see Fujita 2008), future studies should investigate the possibility of differences among ape species in greater detail.

In conclusion, the current data do not neatly fit the predictions made by several hypotheses proposed to explain the looking response here investigated. It appears that positing that subjects knew that they could be wrong is consistent with the results of the three experiments. Although this hypothesis does not deny the contribution of uncertainty in some of the observed responses, it highlights that uncertainty alone cannot be the sole explanation for the observed results. The current findings combined with those based on the escape response method suggest that non-human animals may possess some metacognitive abilities although further work is needed to clarify whether metarepresentation is involved in the kinds of decisions investigated here. At the very least, the current findings should serve to challenge the proponents of the non-

metacognitive account to produce a new set of non-meta-cognitive explanations to account for the current findings.

# References

Barth J, Call J (2006) Tracking the displacement of objects: a series of tasks with Great apes and young children. J Exp Psychol Anim Behav Proc 32:239–252

Basile BM, Hampton RR, Suomi SJ, Murray EA (2009) An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). Anim Cogn 12:169–180

Beran MJ, Smith JD, Coutinho MVC, Couchman JJ, Boomer J (2009) The psychological organization of 'uncertainty' responses and 'middle' responses: a dissociation in capuchin monkeys (*Cebus apella*). J Exp Psychol Anim Behav Proc 35:371–381

Bräuer J, Call J, Tomasello M (2004) Visual perspective taking in dogs (*Canis familiaris*) in the presence of barriers. Appl Anim Behav Sci 88:299–317

Call J (2004) Inferences about the location of food in the great apes (*Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus*). J Comp Psychol 118:232–241

Call J (2005) The self and the other: a missing link in comparative social cognition. In: Terrace HS, Metcalfe J (eds) The missing link in cognition: origins of self-reflective consciousness. Oxford University Press, New York, pp 321–341

Call J, Carpenter M (2001) Do chimpanzees and children know what they have seen? Anim Cogn 4:207–220

Carruthers P (2009) Meta-cognition in animals: a skeptical look. Mind Lang 23:58–89

Crystal JD, Foote AL (2009) Metacognition in animals. Comp Cogn Behav Rev 4:1–16

Foote AL, Crystal JD (2007) Metacognition in the rat. Curr Biol 17:551–555

Fujita K (2009) Metamemory in tufted capuchin monkeys (*Cebus apella*). Anim Cogn 12:575–585

Hampton RR (2001) Rhesus monkeys know when they remember. Proc Natl Acad Sci U S A 98:5359–5362

Hampton RR, Zivin A, Murray EA (2004) Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. Anim Cogn 7:239–254

Inman A, Shettleworth SJ (1999) Detecting metamemory in nonverbal subjects: a test with pigeons. J Exp Psychol Anim Behav Proc 25:389–395

Kornell N, Son L, Terrace H (2007) Transfer of metacognitive skills and hint seeking in monkeys. Psych Sci 18:64–71

Posada S, Colell M (2007) Another gorilla (*Gorilla gorilla gorilla*) recognizes himself in a mirror. Am J Primat 69:576–583

Smith JD (2009) The study of animal metacognition. Trends Cogn Sci 13:389–396

Smith JD, Schull J, Strote J, McGee K, Egnor R, Erb L (1995) The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). J Exper Psychol Gen 124:391–408

Smith JD, Shields WE, Schull J, Washburn DA (1997) The uncertain response in humans and animals. Cognition 62:75–97

Smith JD, Shields WE, Washburn DA (2003) The comparative psychology of uncertainty monitoring and metacognition. Behav Brain Sci 26:317–373

Smith JD, Beran MJ, Redford JS, Washburn DA (2006) Dissociating uncertainty states and reinforcement signals in the comparative study of metacognition. J Exp Psychol Gen 135:282–297

Smith JD, Redford JS, Beran MJ, Washburn DA (in press) Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. Anim Cogn 13 doi: 10.1007/s10071-009-0249-5

Staddon JER, Jozefowiez J, Cerutti D (2007) Metacognition: a problem not a process. PsyCrit 1–5

Suarez SD, Gallup GG (1981) Self-recognition in chimpanzees and orangutans, but not gorillas. J Hum Evol 10:175–188

Suda-King C (2008) Do orangutans (*Pongo pygmaeus*) know when they do not remember? Anim Cogn 11:21–42

Washburn DA, Smith JD, Shields WE (2006) Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. J Exp Psychol Anim Behav Proc 32:85–89