



Remembering Past Exchanges: Apes Fail to Use Social Cues

Amy Lewis^{1,2,3*}, Dorte Berntsen² and Josep Call^{1,3}

¹ School of Psychology and Neuroscience, University of St Andrews, Scotland.

² Center on Autobiographical Memory Research, Aarhus University, Denmark.

³ Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.

*Corresponding author (Email: avml@st-andrews.ac.uk)

Citation – Lewis, A., Berntsen, D., & Call, J. (2018). Remembering past exchanges: Apes fail to use social cues. *Animal Behavior and Cognition*, 5(1), 19-40. <https://doi.org/10.26451/abc.05.01.03.2018>

Abstract – Nonhuman primates can remember events from their distant past. Furthermore, they can distinguish between very similar events by the process of binding. So far, research into long-term memory and binding has focused on the binding of contextual information, such as spatial surroundings. As such, we aimed to investigate if apes can bind and retrieve other types of information, specifically, social information. We presented great apes with three different object types; they learnt to exchange (via reinforcement) one of the object types with one experimenter and another type with a second, different, experimenter. The remaining object type was not reinforced by either of the experimenters. After a delay of two or ten weeks, we assessed the apes' memory of which object type was exchanged with which experimenter. Additionally, we introduced a new experimenter to see if the apes could infer by exclusion that the remaining object type should be exchanged with the new experimenter. The apes successfully remembered which object types were exchanged, but failed to distinguish which object type was exchanged with whom. This failure to bind an object type to a specific person may have resulted from the apes learning to use a rule based on recency, as opposed to learning a conditional rule involving social information. However, results from a second experiment suggested they fail to incorporate social information even when no other information could guide successful performance. Our findings are consistent with research showing long-term memory in primates, but suggest that social information may not be bound in memory as readily as spatial or contextual information.

Keywords – Social cue, Long-term memory, Great apes, Exchange, Binding

In the animal literature, research on memory for past events has shown that information present at the time of encoding can be an effective cue at retrieval (Clayton & Dickinson, 1998; Eacott, Easton, & Zinkivskay, 2005; Martin-Ordas, Berntsen, & Call, 2013; Mendes & Call, 2014). For instance, Eacott et al. (2005) showed that rats could use a contextual feature that was present at encoding and retrieval to successfully locate a desired object. The rats were trained in a two-armed maze to learn that in one context (smooth, black maze) the location of a novel object would be on one arm (e.g., the right), whereas in another context (wire mesh maze) the location would be on the other arm (e.g., left). At test, they could not see the novel object from their starting point, but were able to use the contextual information to successfully determine which arm (where) the novel object was located, thus showing memory of where, and in which context, an item was located. Similarly, Martin-Ordas et al. (2013) found that chimpanzees and orangutans were able to correctly remember the location of a hidden tool when presented with features that overlapped with the tool hiding event. In order to retrieve the correct location, the apes needed to distinguish between similar events that shared many features with each other, such as the same experimenter, the same room, and the use of tools. Only by binding these features together in memory

could the apes distinguish between similar events, as the features in insolation were not diagnostic of a specific event (see also Clayton, Yu, & Dickinson, 2001; Crystal & Smith, 2014).

So far, research investigating cues and binding in animal memory has mainly focused on contextual/environmental information, such as using spatial surroundings to cue memory. In primates, spatial information may be particularly salient due to the need to remember where in the environment food and resources are located. Indeed, evidence suggests that primates are adept at mapping spatial information in the environment to remember food locations (Janmaat, Ban, & Boesch, 2013; Normand, Ban, & Boesch, 2009; Noser & Byrne, 2007). For instance, a study by Menzel (1999) showed that a chimpanzee (Panzee) could recall the hiding location of numerous food items; Panzee successfully recalled each of the hiding locations with great accuracy. Furthermore, she recalled which item was hidden in each location, showing evidence of binding the ‘what’ and ‘where’ features into memory. Consequently, it may be that spatial information is encoded effortlessly, or is easily bound to information in memory, making it effective as a retrieval cue.

The extent to which primates can bind and retrieve other types of information from memory is largely unknown. Social information is particularly relevant in primate societies and is involved in important aspects, such as social hierarchy and reproduction (Cheney, Seyfarth, & Smuts, 1986). Consequently, the ability to recognize other group members is a crucial skill, and has been evidenced in numerous studies (Parr, Winslow, Hopkins, & de Waal, 2000; Pokorny & de Waal, 2009a, b; Rosenfeld & Van Hoesen, 1979). Additionally, primates have been shown to make use of social information in various situations (see Anderson, 1998 for a review). For example, Mineka, Davidson, Cook, and Keir (1984) found that young rhesus monkeys who observed their mother behaving fearfully towards a snake also developed a fear of snakes. Regarding the ability to bind and retrieve social information from memory, a study by Mendes and Call (2014) found some evidence of this. In this study, they found that chimpanzees could successfully remember the locations of food when presented with social cues that overlapped with the food discovery event. During the initial discovery of the food locations, the chimpanzees were released into the enclosure in pairs and an experimenter stood on an observation platform in full view of the subject. At retrieval (24 hours and then 3 months later), the apes were released in the same pairs and the same experimenter stood on the observation platform, thus, the social information present at encoding was also present at retrieval. They found that the chimpanzees searched the hiding locations more often than not, and with greater speed than the initial discovery, suggesting they recalled the locations. However, it is unknown whether the chimpanzees’ memory was due to the presence of their partner, the identity of the experimenter (or any experimenter on the platform), or due to other non-social cues, such as spatial features. As such, it is unclear as to whether the social information was initially encoded, and whether it later acted as a cue at retrieval. Additionally, it has been found that including a social cue during encoding and retrieval of a hiding event does not improve apes’ memory of the event (Lewis, Call, & Berntsen, 2017a, b). Specifically, it was found that the apes’ performance was not better when the same experimenter was present at encoding and retrieval, as opposed to when the experimenter identity differed between encoding and retrieval.

These studies may simply show that primates do not distinguish between one human and another, and are therefore unable to make use of this social information. However, such an explanation is unlikely due to experiments showing that primates are capable of recognizing human faces (Martin-Malivel & Okada, 2007; Sliwa, Duhamel, Pascalis, & Wirth, 2011; Tomonaga, 1999). Furthermore, Schwartz, Colon, Sanchez, Rodriguez, and Evans (2002) found evidence that a gorilla named King could remember which human experimenter previously gave him food. King was tested on his ability to remember what food he had eaten and who had given him the food. To answer these questions, he was supplied with five food cards; four depicted foods that he had not been given (distractors) and one depicted food he had eaten last. Additionally, he was given two name cards on which a name was written in bold. One of the names corresponded to the experimenter whom had given him food most recently, the other was used as a distractor. When asked which food he had eaten last, and who gave him the food, King was able to choose the correct food and experimenter card above chance. Thus, King could distinguish between the two experimenters and remember who had given him what. In a later experiment (Schwartz, Meissner,

Hoffman, Evans, & Frazier, 2004), King was also able recognize which unfamiliar person he had previously seen in a novel event, suggesting that he was able to recognize human faces after a very brief exposure. Although King's performance is consistent with the binding of 'who' information to 'what', it is possible that he could have remembered the information independently; that is, the retrieval of 'what' he ate may have been answered without recalling 'who' gave him the food, and vice versa.

The successful performance by King seems to run counter to the findings of the studies previously mentioned (e.g., Lewis et al., 2017a, b); however, this may be due to the task design. As King was explicitly questioned as to 'who' gave him the food, and received a forced choice option between two people, his attention was explicitly drawn to this social information. In the other studies, this social information was not made explicit and was not questioned directly, and could explain why the social information that was available in these tasks did not lead to successful, or improved, performance. In these cases, the social information may have been outshined (Smith, 1994, 2013; Smith & Vela, 2001). Outshining refers to encoded information that is ineffective as a retrieval cue due to other information providing a stronger cue, for reasons such as cue-overload (Watkins & Watkins, 1975), saliency, or motivational factors. Similarly, social information may also have been overshadowed at encoding. Overshadowing results from the limited capacity to attend and encode information, resulting in salient information being encoded at the expense of less salient information (Smith, 1994, 2013; Smith & Vela, 2001). Thus, it may be that the social information was overshadowed during encoding, or outshined at retrieval, by the presence of other more salient information.

As such, it is unclear as to whether primates can bind and retrieve social information from memory, particularly when other salient information is available. We therefore designed a paradigm whereby apes could only correctly discriminate between alternative choices by using social information, enabling us to see if, in the absence of other relevant information, they could bind and retrieve social information. As previous research has shown that primates are capable of exchanging objects for rewards (Osvath & Persson, 2013; Pele, Dufour, Thierry, & Call, 2009), and of succeeding in discrimination tasks (Anderson, 1996; Itoh, Izumi, & Kojima, 2001), we chose to use an object exchange discrimination task. Apes were presented with three different types of objects, and were trained via reinforcement to exchange one of the object types with the first experimenter, and one of the other types with a second experimenter; a third type was not trained. After a delay, apes were assessed on whether they could remember which object type was rewarded with which experimenter. We used a retention period of two and ten weeks (between-subjects), as previous research has shown that apes are capable of remembering information after both these time periods (Lewis et al., 2017b; Martin-Ordas et al., 2013; Mendes & Call, 2014), with no significant decline in performance from two to ten weeks (Lewis et al., 2017b). Furthermore, if the apes show successful binding of a particular object type to a specific person, we wanted to investigate whether they could use this information to infer by exclusion. Specifically, if they learn that object type A is bound to Experimenter 1, and object type B is bound to Experimenter 2, would they assume that object type C should be bound to Experimenter 3? We know that primates and other species are able to infer by exclusion using causal and spatial-temporal information (see Beran & Washburn, 2002; Völter & Call, 2017), but it is less clear if they can do so with more abstract information, such as social information.

Experiment 1

Method

Subjects. Twenty-five apes took part in this experiment (see Table 1). Subjects consisted of 12 chimpanzees (*Pan troglodytes*), seven bonobos (*Pan paniscus*), and six orangutans (*Pongo abelii*), aged between 7 – 42 years ($M = 19.32$). Subjects were tested between August 2015 and April 2016.

Apparatus. The objects to be exchanged were wooden painted blocks. Three different types were used: a green rectangle (H: 4.9 cm X L: 2.3 cm X D: 1.2 cm), a blue cube (H: 2.5 cm X L: 2.5 cm X D: 2.5 cm), and a pink triangle (H: 2.1 cm X L: 4.5 cm X D: 2.9 cm; see Figure 1). There were four objects of each type, resulting in a total of 12 objects. Objects were replaced when they became damaged (broken

or dis-colored objects). The objects were given to the subject either via a Plexiglas transparent container attached to the meshing (bonobos), or via a food hatch (chimpanzees and orangutans). All objects were exchanged through meshing (see Figure 2 for the experimental set-up). An opaque panel was used in later trials, which served to block the apes' view of the experimenter, and vice versa.

Table 1

Participation of Subjects by Experiment.

Subject	Species	Sex	Age (years)	Experiment 1	Experiment 2
Fimi	Bonobo	Female	7	✓	-
Gemena	Bonobo	Female	10	✓	-
Luiza	Bonobo	Female	11	✓	-
Lexi	Bonobo	Female	16	✓	-
Yasa	Bonobo	Female	18	✓	-
Kuno	Bonobo	Male	19	✓	-
Jasango	Bonobo	Male	25	✓	-
Bangolo	Chimpanzee	Male	7	✓	-
Kofi	Chimpanzee	Male	11	✓	-
Lobo	Chimpanzee	Male	12	✓	-
Tai	Chimpanzee	Female	13	✓	-
Lome	Chimpanzee	Male	14	✓	-
Alex	Chimpanzee	Male	15	✓	✓
Annett	Chimpanzee	Female	16	✓	-
Frodo	Chimpanzee	Male	22	✓	-
Jahaga	Chimpanzee	Female	23	✓	-
Zira	Chimpanzee	Female	20	-	✓
Hope	Chimpanzee	Female	26	-	✓
Dorien	Chimpanzee	Female	35	✓	-
Fraukje	Chimpanzee	Female	40	✓	-
Frederike	Chimpanzee	Female	42	✓	✓
Suaq	Orangutan	Male	7	✓	✓
Raja	Orangutan	Female	12	✓	-
Padana	Orangutan	Female	18	✓	✓
Dokana	Orangutan	Female	27	✓	✓
Pini	Orangutan	Female	28	✓	✓
Bimbo	Orangutan	Male	35	✓	-

Note: Ticks indicate subject participation and dashes indicate absence of participation. Age shown is the age of the subject at the time of their first Experiment.

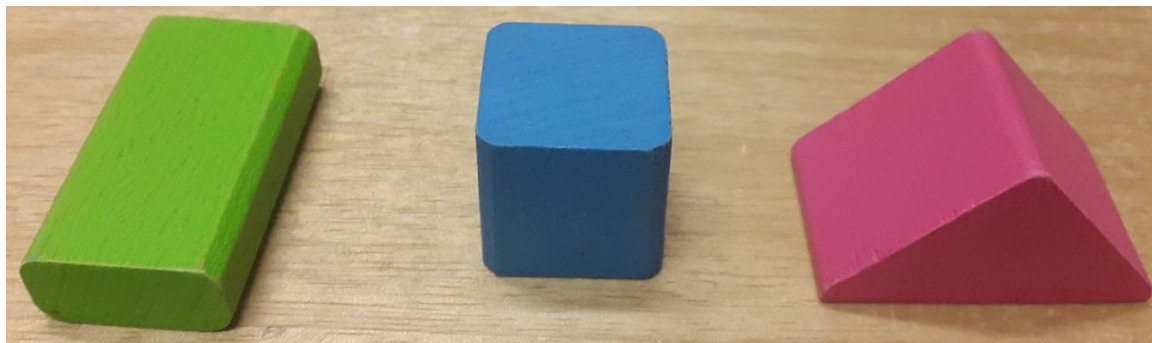


Figure 1. The three object types used in Experiment 1.



Figure 2. The experimental set-up. Objects were placed into the food hatch at the start of each trial (pictured to the left of the ape). Objects were exchanged through the lower mesh. Photo printed with permission from the Max Planck Institute of Evolutionary Anthropology.

Design and procedure. There were three different experimenters (E1, E2, E3) and three types of objects (see apparatus). The three experimenters were all familiar to the apes, and had tested the apes on at least one other study prior to this experiment. Each object type was assigned a letter (A, B or C); for instance, the green rectangle was assigned A, the blue square was assigned B, and the pink triangle was assigned C (assignment was counterbalanced between-subjects). At training, E1 rewarded object type A, and E2 rewarded object type B. The subjects trained with E1 and E2 only, thus they were never

introduced to E3 and object type C was never rewarded. They trained with one experimenter at a time, and did not begin training with the second experimenter until passing training with the first. The experimenter that they trained with first was randomized between-subjects (i.e. some began with E1, and some with E2).

There were four trials per session, with one session a day. To pass a session the subject had to exchange the correct object a minimum of 13 (out of 16) times in the first four exchanges of each of the four trials (81.25% correct); that is, only the first four exchanges per trial were counted, and they had to exchange a minimum of 13 correct objects across the four trials to pass the session. Once subjects had successfully passed two sessions, an opaque panel was inserted into the panel frame above the mesh. This blocked the ape's view of the experimenter (and vice versa), ensuring that the experimenter did not unintentionally cue the subject as to which object they should exchange. All subjects needed to pass six training sessions (two transparent and four opaque), with both E1 and E2.

Once apes had passed these six sessions, we checked whether they exchanged the correct object type at least 50% of the time in the first four exchanges of the four opaque passed sessions. As there were sixteen correct exchanges per session, the apes may have learnt which object type was correct *during* a session, rather than learning over sessions which object type was correct. If this were the case, we would expect to see poor performance in the first few exchanges of a session, whereby the apes exchanged the objects randomly until they learned which one was correct. Therefore, we checked this was not the case by examining whether they exchanged the correct object the majority (at least 50%) of the time in the first four exchanges of the four opaque sessions. If subjects failed to do this, they received additional training sessions until they reached a 50% (or greater) exchange rate in the last four opaque sessions. If subjects failed to complete two consecutive sessions, or did not pass at least one session by the 15th session, they were dropped from the study. A session was cancelled and re-scheduled if the ape did not respond after five minutes; this happened with Frodo (six times), Fraukje (once), and Kuno (twice).

After successful completion of training, subjects completed a test. The test took place either two weeks (10-20 days, $M = 15$), or ten weeks (68-95 days, $M = 76$) after completion of training (between-subjects). There were three test sessions (within-subjects), one with each experimenter (E1, E2, E3), and each test was conducted on separate, consecutive days (with the exception of Fimi, who had a two-day delay between test one and test two). The order of the three test sessions was randomized between-subjects (see Figure 3 for an overview of the training and test phases). During each test session, subjects completed four trials, and were rewarded for passing any object through the meshing (non-differential reinforcement). This was to prevent further learning during the test, ensuring that we only assessed the subject's memory of what they had previously learnt. A trial finished when all four correct objects had been exchanged; a correct object was an object of the type corresponding to the present experimenter (e.g., object type A with E1). The opaque panel was used in all sessions to prevent any possible unintentional cues from the experimenter. If the apes successfully learned to bind an object type with a specific experimenter during training, they should preferentially give the correct objects to the correct experimenters at test (e.g., during E1's test the subject should give object type A). If they were able to bind and recall this information perfectly, it would take them four exchanges to give all the correct objects. If the apes were able to show inference by exclusion, in E3's test they should give object type C. Additionally, object type C served as a control object in E1 and E2's test, whereby if the apes failed to remember that either type A and B objects were rewarded, then they should not exchange them preferentially over type C. All sessions were videotaped (with the exception of one of Bimbo's and Alex's sessions due to a camera error), and all subjects were tested individually, except for mother and infants, who were not separated ($N = 4$).

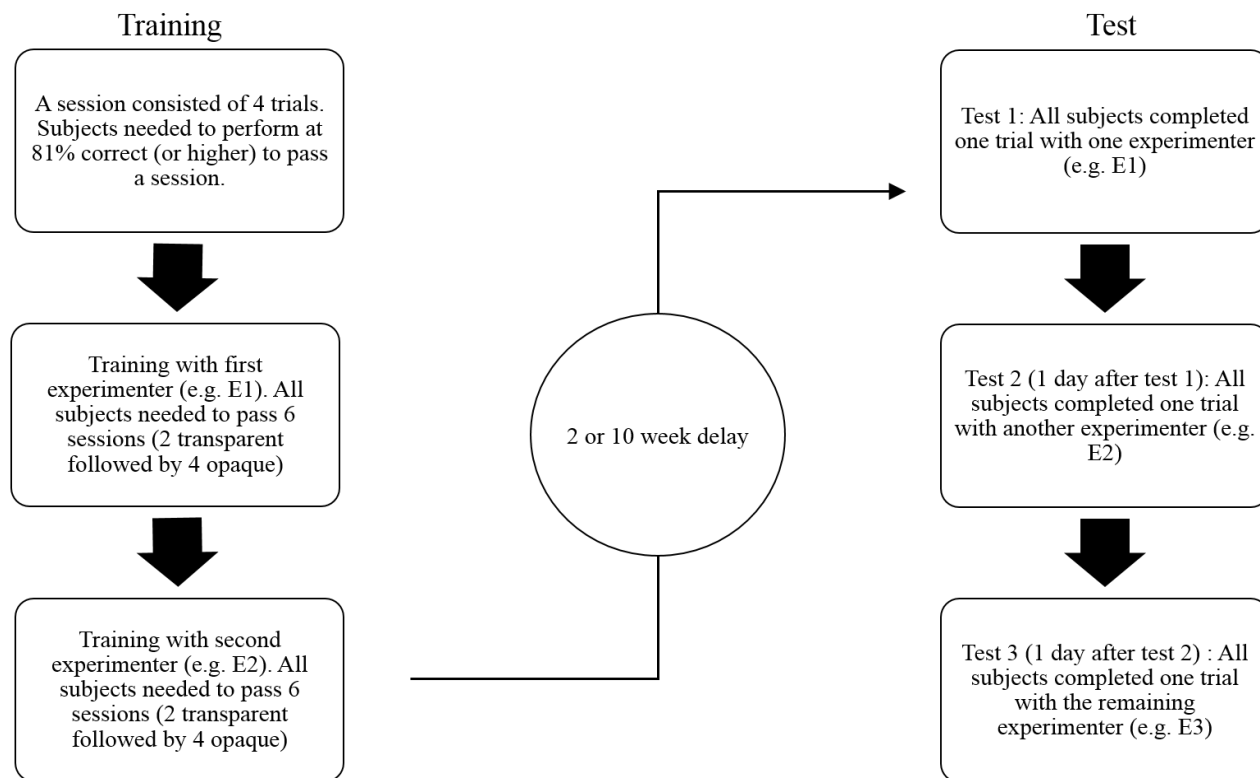


Figure 3. Example order of training and test phases.

Training procedure. At the start of a session, E1 put all twelve objects into the food hatch (or container). The apes were able to access all the objects from here. E1 then sat in front of the meshing and asked the subject for an object by holding one palm face up and a reward (grape) in their other hand. If the subject did not pass an object through the mesh, E1 verbally encouraged them. If they still did not respond after five minutes, the session was cancelled and re-scheduled. When the subject gave a correct object (i.e., object A), E1 immediately rewarded them with a grape and then placed the object into a bucket to the left (here-after object bucket). If the object was incorrect (i.e., type B or C), no reward was given and the object was placed into the object bucket. In some cases, the infant took the objects or stole the food reward. If an object could not be retrieved from an infant, an additional object (of the same type) was put into the container/food hatch. There were four possible correct objects per trial, when all four correct objects had been exchanged, the trial ended. A new trial was initiated by E1 putting the contents of the object bucket into the container/food hatch. The procedure above describes training with E1, training with E2 differed only in that E2 was the experimenter, and the correct object type was B and the incorrect object types were A and C.

Test procedure. As with training, the experimenter put all twelve objects into the container/food hatch and sat in-front of the meshing. The exchanging followed the same procedure as training, except all exchanges were rewarded, regardless of whether the object was correct or not. A trial finished once all four correct objects had been exchanged, the correct object was the type that corresponded to the present experimenter (e.g., type A with E1; see Design).

Data analysis. All analyses were conducted using SPSS version 20.0.0, with the alpha level set to 0.05 (unless otherwise specified). Greenhouse-Geisser was reported when Mauchly's test of sphericity was significant.

Training. During training, the dependent variable was the number of sessions taken to reach criterion (minimum six). We were interested in whether learning rates differed between species, and whether the apes' performance with the second experimenter was influenced by training with the first

experimenter. This is because the apes may show difficulties with inhibiting the object type that was rewarded with the first experimenter when they begin training with the second experimenter, as has been observed in other studies on inhibition in apes (Beran & Evans, 2009; Uher & Call, 2008; Vlamings, Hare, & Call, 2010; Vlamings, Uher, & Call, 2006). Differences in performance between the first and second experimenter may also be due to proactive interference, whereby the apes have difficulties learning new information as a result of previously learnt information (Anderson & Neely, 1996) As such, we conducted a 3 (species) X 2 (experimenter; first and second) mixed ANOVA.

Test. Each subject took part in three tests, one with each experimenter (E1, E2, E3). If the apes successfully bound and recalled which object type was rewarded by which experimenter, then performance in tests with E1 and E2 should be better than E3 (as E3 was never present at training). However, if the apes also performed well in the test with E3, this would indicate successful inference by exclusion. We tested for any differences between the 3 tests (within-subjects), between species and between delay groups (between-subjects), using a 3 (test) X 3 (species) X 2 (delay) mixed ANOVA. The dependent variable was the number of exchanges needed to exchange all four correct objects (minimum 4, maximum 12); a low score indicated higher levels of performance. To further test whether the apes were able to successfully bind object type A to E1 and object type B to E2, we ran a 2 (test; E1, E2) X 2 (object type; A, B) repeated measures ANOVA, with the dependent variable as the number of exchanges. Only the first four exchanges with each experimenter were counted, as there were four correct objects per experimenter. If the subjects bound E1 to object type A and E2 to object type B, we should see an interaction whereby the apes exchange more A type objects with E1 than E2 and more B type objects with E2 than E1.

Additionally, we were interested in whether performance in tests with E1 and E2 was influenced by the experimenter with whom they last trained, independent of the delay period. This is because the apes may show a preference for the most recently rewarded object. As such, we tested for this using a 2 (test; E1, E2) X 2 (experimenter last trained) X 2 (delay) mixed ANOVA; the dependent variable was the number of exchanges needed to exchange all four correct objects (as with the first ANOVA described). Test was within-subjects and experimenter last trained and delay were between-subjects. Alpha level for these analyses were adjusted to 0.016 to correct for multiple testing (Bonferroni correction).

Reduced sample. The analyses described above were conducted on the data from all subjects. Additional analyses similar to the ones conducted for the full sample were also conducted on a reduced data set that excluded three orangutans (Bimbo, Suaq and Padana). These orangutans had previously participated in a pilot task similar to Experiment 1, as such, the additional analyses enabled us to establish whether experience in the pilot task may have influenced performance in Experiment 1. These tests replicated the reported for the full sample and will not be described further (Table 2 includes the descriptive statistics for both the full and reduced sample).

Table 2

<i>Mean Number of Sessions Needed to Complete Training with the First and Second Experimenter by Species</i>				
		Bonobo	Chimpanzee	Orangutan
First experimenter	Mean	10.86 (10.86)	11.38 (11.38)	9.40 (9.00)
	SD	2.61 (2.61)	2.50 (2.50)	1.67 (1.41)
Second Experimenter	Mean	17.14 (17.14)	11.25 (11.25)	12.00 (12.50)
	SD	3.58 (3.58)	2.05 (2.05)	3.54 (2.12)

Note: Numbers in parenthesis refer to the sample excluding three orangutans (reduced sample).

Results

Training. One chimpanzee was dropped during training with the first experimenter (Dorien) and three with the second experimenter (Annett, Tai, Fraukje) due to failing 15 sessions. Additionally, one orangutan (Raja) was dropped with the second experimenter after failing to complete two consecutive sessions, resulting in a total of 20 apes. Of the subjects that passed six sessions, all exchanged the correct object type in the first four exchanges of the four opaque sessions more than 50% of the time ($M = 87.50\%$, $SD = 14.44\%$), showing that they retained what they had learnt across sessions. The means and standard deviations by species and experimenter can be seen in Table 2. We found a main effect of experimenter, $F(1, 17) = 26.94$, $p < .001$, $\eta_p^2 = .613$, whereby subjects took longer to learn with the second experimenter than the first. This effect was due to the apes continuing to exchange the rewarded object type from the first experimenter in the first few sessions with the second experimenter (see Table 3). No main effect of species was found, but a trend was observed which suggested that bonobos took longer to complete training than the chimpanzees and orangutans, $F(2, 17) = 3.39$, $p = 0.058$, $\eta_p^2 = .285$. Additionally, we found an interaction between species and experimenter, $F(2, 17) = 12.62$, $p < .001$, $\eta_p^2 = .598$, which showed that bonobos and orangutans took more sessions to complete training with the second than the first experimenter, but that chimpanzees did not (and were marginally better with the second). However, these results should be taken with caution; 3 of the chimpanzees failed to complete training with the second experimenter after 15 sessions and were dropped from the study (Annett, Tai, Fraukje), thus their data were not included. The failure to pass training shows the difficulty with learning with the second experimenter, and thus there is likely to be no real differences between chimpanzee learning rates as compared to the orangutans and bonobos.

Table 3

Mean Number of Times the Correct Object Type from the First Experimenter was Given to the Second Experimenter (Percentages in Parentheses) in the First Three Sessions

		Bonobo	Chimpanzee	Orangutan
Session 1	Mean	12.29 (76.81%)	13.50 (84.38%)	13.60 (85.00%)
	SD	2.69	2.19	3.21
Session 2	Mean	11.14 (69.63%)	12.38 (77.38%)	10.20 (63.75%)
	SD	2.67	2.62	4.60
Session 3	Mean	8.86 (55.38%)	8.88 (55.50%)	7.20 (45.00%)
	SD	3.29	3.91	5.26

Note: Only the first 4 exchanges per trial were counted, with 4 trials per session (total of 16 exchanges).

Test. One chimpanzee (Frederike) was dropped at test due to experimental error, resulting in a total of 19 subjects. We found a main effect of test, $F(1.15, 14.95) = 16.77$, $p = .001$, $\eta_p^2 = .563$, but no effects of species or delay ($F_s < 0.71$, $p_s > .51$). None of the interactions between species, test and delay were significant ($F_s < 2.73$, $p_s > 0.10$). The main effect of test showed that the apes needed more exchanges to reach four correct objects with E3 (i.e., C type objects; $M = 11.42$, $SD = 1.26$) than they did in tests with E1 ($M = 7.42$, $SD = 2.60$) and E2 ($M = 6.89$, $SD = 2.42$); that is, A and B type objects respectively (see Figure 4). The greater number of exchanges in the test with E3 suggests subjects were avoiding the non-reinforced object and preferentially selecting the reinforced objects. Additionally, tests with E1 and E2 were not significantly different $t(18) = 0.49$, $p = .63$, meaning subjects took the same number of exchanges to exchange the correct object in both tests. Although performance was better in tests with E1 and E2 compared to E3, further analysis revealed that the apes were not successfully binding

object type A with E1 and object type B with E2, as shown by a lack of interaction between experimenter and object type, $F(1, 18) < 0.001$, $p = 1$; that is, the apes were not exchanging object type A more often with E1 ($M = 1.47$, $SD = 1.74$) compared to E2 ($M = 1.47$, $SD = 1.35$), nor type B more often with E2 ($M = 2.47$, $SD = 1.35$) compared to E1 ($M = 2.47$, $SD = 1.74$).

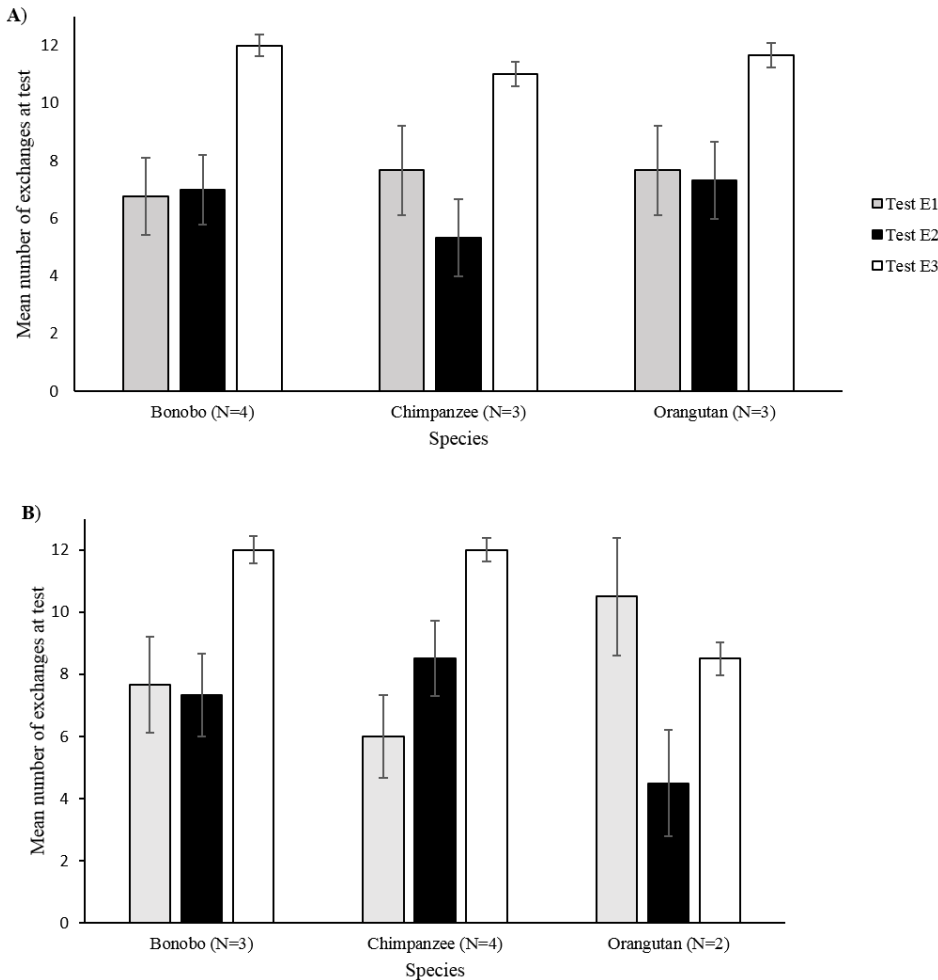


Figure 4. Mean number of exchanges needed to exchange all 4 correct objects in each test by species. A) depicts subjects in the 2 week delay period, B) depicts subjects in the 10 week delay period. Error bars denote the standard error.

An additional analysis considering only tests with E1 and E2 revealed a significant interaction between test (i.e., between A and B type objects) and the experimenter with whom they trained last $F(1, 15) = 17.35$, $p = .001$, $\eta_p^2 = .536$. There was no 3-way interaction when delay was included, $F(1, 15) = .01$, $p = .93$. This meant that subjects performed better (had fewer exchanges) in the test in which the correct object type was the type last trained with, independent of delay (see Figure 5). For example, if they trained with E2 (object type B) last, they performed better in the test with E2 (in which object type B was correct) than the test with E1 (in which object type A was correct). Thus, subjects preferentially exchanged the object type that was reinforced last. As subjects took longer to train with the object that was reinforced last (i.e., training data showed greater mean number of sessions with the second experimenter, as mentioned previously, *cf.* Table 2), we tested whether this object type was preferred due to being reinforced more; a non-significant correlation between the number of training sessions with the last experimenter and the number of exchanges on the corresponding test showed that apes preferred the last rewarded object type regardless of how much it had been reinforced; $r(19) = -.09$, $p = .72$.

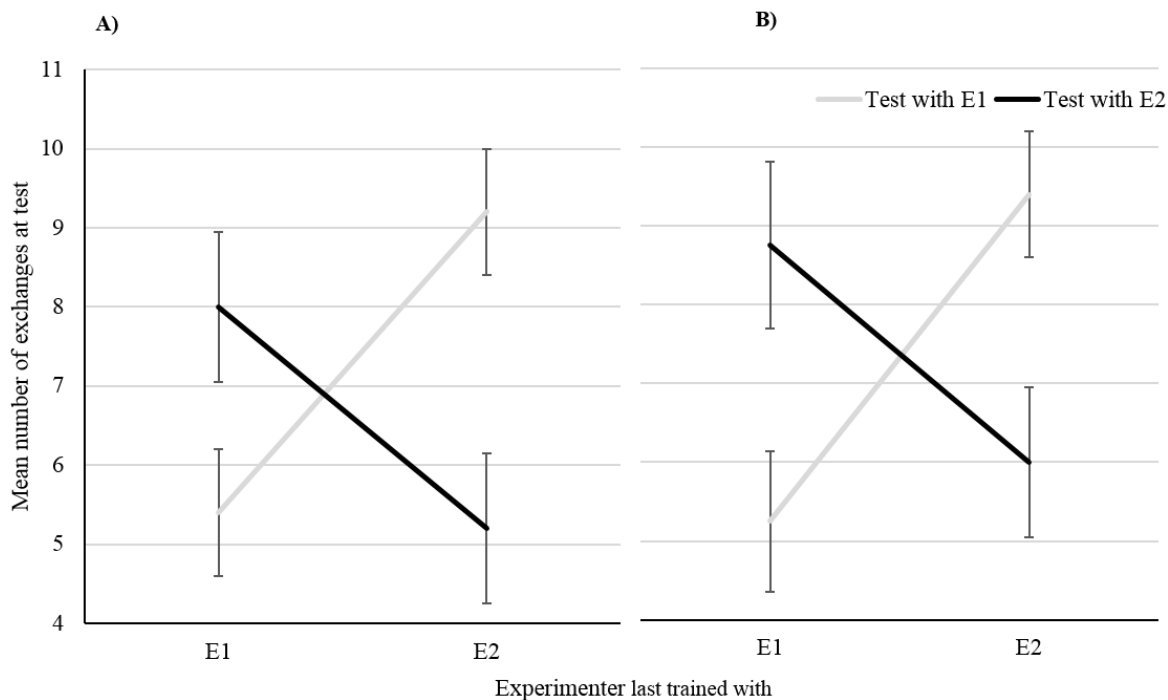


Figure 5. Mean number of exchanges in the test with E1 and E2 by experimenter with whom they last trained. A) Depicts subjects in the 2-week delay period, B) depicts subjects in the 10-week delay period. Error bars denote standard error.

As the apes were non-differentially rewarded at test they may have adapted their response based on this new contingency, rather on what they had previously learned during training. As such, we conducted additional analyses using only the first exchange from each of the three tests. To test for differences between the three tests, we compared the number of correct and incorrect first exchanges between the three tests using an exact Cochran's Q test. Results indicated that performance in the three tests was significantly different ($Q(2) = 10.33, p = .005$), with worse performance in test with E3 (0% correct exchanges) than tests with E1 and E2 (37% and 58% correct first exchanges respectively). An exact McNemar test revealed no difference between performance in tests with E1 and E2 ($p = .48$). These results replicate the main analysis, with subjects performing significantly worse in the test with E3 compared to tests with E1 and E2, and with comparable performance between tests with E1 and E2.

As with the main analysis, we also tested whether performance in tests with E1 and E2 was influenced by the experimenter with whom they last trained. We calculated whether the number of apes that trained with E1 last and gave object A in the first exchange of both tests was above chance, and likewise if the number of apes that trained with E2 last and gave object B in both tests was above chance. Chance was calculated as 0.33, as three possible object types could be exchanged first, and alpha level was corrected to 0.025. Binomial tests revealed that of the apes that trained with E1 last ($n = 9$), 78% gave object A in the test with E1, and 89% gave A in the test with E2 ($p < .001$); in both cases this was significantly above chance ($p < .01$). Of those who trained with E2 last, all ten subjects gave object B in both tests, which was significantly greater than chance ($p < .001$). Furthermore, 79% of all subjects (15 of 19) first exchanged the object type that was rewarded last at training across all three tests. These results are consistent with the main analysis, in which apes preferentially exchanged the object type that belonged to the experimenter they last trained with, regardless of test condition.

Discussion

We found that the apes successfully remembered which object types were rewarded two and ten

weeks previously, but found no evidence that they bound the object type to the experimenter. Subsequently, they did not infer that object type C belonged to E3. This result is open to multiple interpretations, however, we believe that forgetting is not one of them, in part since we saw no effect of delay. The apes' performance in E1 and E2's test was significantly better than performance in E3's test, indicating that the apes preferentially gave object types A and B before type C, clearly demonstrating that they remembered which object types had been rewarded at training whilst avoiding the type that had never been rewarded. This is consistent with other studies that show apes can learn and remember abstract associations between objects and rewards (Osvath & Persson, 2013; Vlamings et al., 2006). In the present study, this effect was independent of having a two-versus ten-week delay.

Additionally, the apes preferentially gave the object type that belonged to the experimenter with whom they trained with last, suggestive of a dominance of the most recently rewarded object. Again, this effect was independent of delay. Furthermore, this effect was not explained by an increase in number of training sessions with the last experimenter (i.e. more reinforcement); apes' performance on the test that corresponded to the last experimenter they trained with did not correlate with the number of training sessions they had (e.g., a greater number of sessions with E1 did not correlate with fewer exchanges in the test with E1).

One likely explanation for the apes failing to bind information to the identity of the experimenter may be a failure to interpret the task in the way we intended. In order to succeed, the apes needed to bind information about which experimenter was present to which object type was rewarded. Our training procedure may not have made this clear to the apes. We trained the apes with each experimenter in blocks, only once they had completed training with the first experimenter did they begin with the second. As such, to pass training the apes may simply have learnt that only one object type is rewarded, and that the rewarded object type arbitrarily changed once during training, without interpreting the identity of experimenter as the signal of the change in rewarded object type. Such an interpretation is supported by the finding that the last rewarded object type was the preferred object type; if the apes believe that the rewarded object type changed, as opposed to being dependent upon the present experimenter, they should preferentially give the last rewarded object type. To fully ensure that the apes had understood the nature of the task, training with both experimenters needed to be run in parallel, meaning that the correct object type would have changed between sessions. Experiment 2 was designed to meet this aim.

The poor performance with E3 is likely due to the failure to infer by exclusion that object C should be given to E3. If the apes failed to learn and remember the relationship between object type A and E1 and object type B and E2, it is not possible to infer by exclusion that the remaining object must go with E3. However, we also acknowledge that the apes may have failed to make this inference even if binding was successful, and instead may have chosen to avoid object C due to it never being reinforced.

With regards to our other findings, we found a trend in which bonobos took longer to complete training than the chimpanzees and orangutans. This was mainly driven by the high average number of sessions needed to complete training with E2 (see Table 2). Such performance may reflect increased difficulty for bonobos with inhibiting a previously successful action. These findings can be seen as consistent with Wobber, Wrangham, and Hare (2010) who found that bonobos were just as quick as chimpanzees to learn a discrimination rule, but were slower to learn the reversal rule, in which the incorrect and correct response swapped; this was particularly true for younger bonobos. However, the findings of this current experiment should be taken with caution, due to the non-significant trend and the relatively small sample size of bonobos ($N = 7$). Additionally, although chimpanzees were significantly quicker at completing training with the second experimenter compared to the other species, these results are unlikely to reflect any true species differences due to three chimpanzees who failed training with the second experimenter and were removed from the analysis.

Regarding delay, no difference was found between the two and ten week conditions, thus all subjects showed long-term memory of reinforced objects for up to ten weeks. This adds to the growing literature on long-term memory in primates, which has already shown primates ability to remember events and spatial locations (Kano & Hirata, 2015; Lewis et al., 2017a, b; MacDonald, 1994; Martin-Ordas et al., 2013; Mendes & Call, 2014; Menzel, 1999), to recognize people and places (Murai, Tanaka,

Tomonaga, & Sakagami, 2011), and to retain knowledge over long periods of time (Beran, 2004; Beran & Heimbauer, 2015; Fagot & Cook, 2006; Laska, Alicke, & Hudson, 1996; Martin-Ordas et al., 2013; Vale et al., 2016).

Experiment 2

The purpose of this second Experiment was to resolve the recency issue that likely led to successful performance during training in Experiment 1. In Experiment 1, the apes may have learnt a reversal discrimination, in which at first one object type was rewarded (e.g., A) and then a previously unrewarded object type was rewarded instead (e.g., B). At test, they then preferentially exchanged the last rewarded object type (a recency effect). As such, this experiment ensured successful performance could only be achieved by learning that the rewarded object is dependent upon the identity of the experimenter.

The design remained very similar to Experiment 1, except for three important changes. First, no object type C was used. The purpose of the C object in the previous Experiment was to see whether the apes would remember at test which object types were reinforced and which types were not, and to see if the apes could infer by exclusion. The data clearly showed that the apes remembered the rewarded object types, thus we did not need to establish this again, and as we failed to show evidence of binding the social information in Experiment 1, the additional complex element of inference by exclusion was removed here. Second, training with E1 and E2 was randomized between sessions. This meant the apes could not initially learn that one object type was always rewarded, and that this then switched once during training. Neither could they learn a new rule, such as the rewarded object type changed each session, as the randomization of sessions meant that there was no pattern or order to the sessions. As such, in order to determine which object type was rewarded in a session the apes needed to attend to the identity of the experimenter. Third, we did not include a test. The test in the previous experiment enabled us to see if the apes could recall what they had learnt after a long delay. Although they did not learn what we intended them to learn, they did remember which objects are rewarded after both two and ten weeks delay, thus we did not need to test for this again. In short, the sole aim of Experiment 2 was to see whether the apes could learn which object should be exchanged with whom.

Method

Subjects. Four chimpanzees (aged 16–43, $M = 26.25$) and four orangutans (aged 8–29, $M = 21$). All but two of the chimpanzees had participated in Experiment 1 (see Table 1). All subjects were tested during September 2017 (it was not possible to test the bonobos during this time, and thus no bonobos were included in this experiment).

Apparatus. Two types of wooden objects were used; an orange rectangle and a yellow cube (the same dimensions as Experiment 1; see Figure 6). There were four objects of each type, resulting in a total of eight objects. All other apparatus remained the same as Experiment 1.

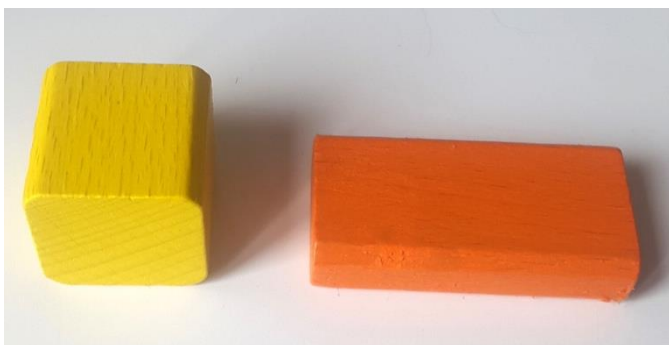


Figure 6. The two object types used in Experiment 2.

Design and procedure. The design remained the same as Experiment 1, except for the following important changes; there were only two types of objects (4 X A and 4 X B), training with E1 and E2 was randomized between sessions (with the stipulation that neither experimenter conducted more than two consecutive sessions), and there was no test (for reasoning behind the changes, see above). All subjects received the same order of sessions, except for cases in which a subject did not want to participate or could not participate on a given day. In these cases, sessions were rearranged for another day (this occurred four times with Suaq, twice with Dokana and once with Pini and Padana). All subjects began their first session with E2. Additionally, E1 was the same Experimenter as in Experiment 1, however E2 changed. This was due to the former E2 from Experiment 1 being unable to participate. The E2 in this current experiment had conducted one other test with the apes previously. Although this meant that E1 was more familiar to the apes than E2, and had participated in a previous exchange task (Experiment 1), the results suggest that performance as a group did not differ between E1 and E2 (see Results).

The criteria to pass a session remained the same as Experiment 1, and as in Experiment 1, all subjects needed to pass six sessions (2 transparent, 4 opaque) to pass training. However, the number of sessions the subjects completed changed slightly. All subjects received a minimum of eleven sessions with each experimenter. Once eleven sessions had been completed, subjects that had passed at least one of the eleven sessions with each experimenter received additional sessions to enable them to have the opportunity to pass training (the minimum number of sessions was set to eleven, because by session eleven with E1 and E2 in Experiment 1, all subjects participating in this current experiment had passed at least two sessions ($M = 5.67$, $SD = 1.16$), and on average had completed training within ten sessions (range 8-14; $M = 9.67$, $SD = 1.78$), with the exception of Zira and Hope who did not take part in Experiment 1). The number of additional sessions was dependent upon performance; for every additional session a subject passed with E1 and E2, they received another session each with E1 and E2 until six sessions were passed with each experimenter. If they failed one of the additional sessions with either experimenter, training ended and it was counted as failed (e.g., if they passed session 12 with E1 but failed with E2, training was failed). If subjects passed training with one experimenter before the other, training still continued with both experimenters until the subject either passed with both experimenters or failed training.

As with Experiment 1, once six sessions had been passed we checked whether the correct object type was exchanged the majority of the time in the first four exchanges of the four passed opaque sessions. As there were only two object types, the correct object needed to be exchanged on average at least three of four times in the first four exchanges (75% of exchanges). If subjects were exchanging randomly, we would expect approximately two of four exchanges to be correct in the first trial of each session (50% of exchanges). Failure to exchange correctly at least 75% of the time may reflect performance based on learning which object was correct *during* a session, as opposed to knowing which object was correct at the start of the session, which would be the case if they had successfully learned the discrimination rule. Subjects that did not exchange the correct object at least 75% of the time received additional training sessions until they reached 75% correct exchanges in the last four opaque sessions.

Failure to pass training would indicate difficulty with binding ‘who’ to ‘what’, whereas passing training would suggest the apes are able to successfully do this when there is no other information to guide successful performance (such as a reversal rule in Experiment 1).

The procedure remained the same as Experiment 1, except only eight objects were present (4 X A and 4 X B) and there were no test sessions.

Data analysis. The dependent variable was fail or pass (i.e., subjects either passed training or did not). We compared whether the number of subjects that failed and passed was significantly different to chance, using a two-tailed exact binomial test. As there were only two possible outcomes (fail or pass), chance was set to 0.5. We also assessed whether performance improved over time, by comparing whether the average number of correct exchanges in the first four exchanges of the four trials of a session (max. number of correct exchanges = 16) was greater in the last five sessions, compared to the first five sessions. For each subject, we calculated a single score based on the average of the combined scores of

the first five sessions with E1 and E2 (first 5 score), and an average of the combined scores of the last five sessions with E1 and E2 (last 5 score). We then compared the first 5 and last 5 scores using a paired t-test.

To assess whether the apes' performance differed between E1 and E2 (as a result of previous experience with E1), we compared the average number of correct exchanges in the first four exchanges of the four trials of a session (max. number of correct exchanges = 16) between E1 (E1 score) and E2 (E2 score). A single score for each subject was calculated as the average score across all completed sessions (e.g., if the subject completed 12 sessions with E1, the average score across those 12 sessions was the E1 score). We then compared the E1 scores and E2 score using a paired t-test.

Results

None of the eight subjects successfully passed training; binomial tests revealed this performance was significantly worse than chance ($p = .008$). However, three of the apes successfully passed at least one session with both E1 and E2, but then subsequently failed to pass session 12 with both experimenters (see Figure 7B,E-F). Performance between the first 5 sessions ($M = 9.16$, $SD = 1.95$) and last 5 sessions ($M = 9.93$, $SD = 1.35$) did not differ, $t(7) = 1.20$, $p = .27$, suggesting the apes' performance did not improve over time, and performance between E1 ($M = 9.73$, $SD = 1.82$) and E2 ($M = 9.20$, $SD = 1.47$) did not differ, $t(7) = .89$, $p = .40$, suggesting that previous experience with E1 did not influence performance.

On an individual level, the performance of the apes was quite varied. Three of the apes (Pini, Hope, and Dokana) exchanged the objects randomly with both E1 and E2, performing roughly at chance levels and showing little improvement over time. As such, they failed to pass a single session with either experimenter (Figure 7D, G-H). Zira showed a similar pattern, until the last three sessions when her performance improved with E1 but deteriorated with E2 (Figure 7C), suggesting she preferentially learned to give object type A to both experimenters. Alex's performance with E2 seemed to improve over time, but at the detriment to performance with E1. However, in the last two sessions his performance with E1 improved whilst his performance with E2 remained good, suggesting he may have begun to understand the conditional discrimination rule. As a result, he was given an additional session each with E1 and E2 (despite not passing a session with E1) to see whether his performance would continue to improve; however, he failed this session with E2 (Figure 7A). Frederike and Padana passed multiple sessions with both E1 and E2, but failed to reach the criteria of six passed sessions with each experimenter, and failed to pass session 12 with E2 and E1 respectively, suggesting they had not learnt the conditional discrimination rule (Figure 7B, E). Suaq successfully passed six sessions each with E1 and E2; however, he did not exchange the correct object type above 75% in the first four exchanges of his last four opaque sessions with either experimenter. He instead reached only 68.75% correct, before subsequently failing to pass session 12 with E2, suggesting he too had not learnt the conditional discrimination rule (Figure 7F).

Discussion

The results show that all of the apes failed to pass training, meaning that they did not learn to exchange one object type with one experimenter, and the other object type with the other experimenter. Furthermore, their performance did not improve over time, as shown by no difference in the number of correct exchanges between the first and last five sessions. These results suggest that given more sessions, the apes would still have failed to learn that the correct object type was dependent upon the experimenter present. However, as performance did numerically improve over time, we acknowledge that a lack of power may also explain this null effect.

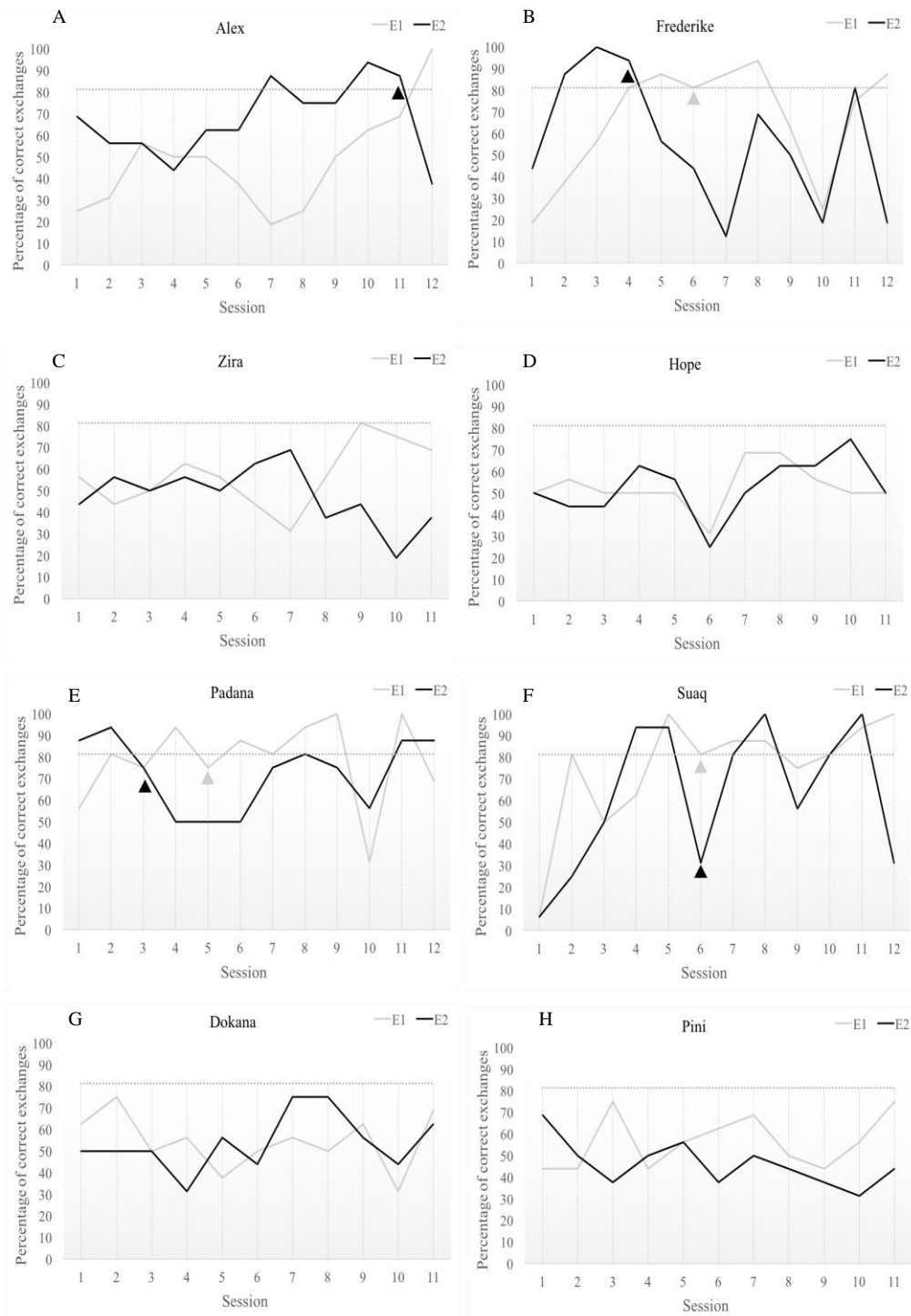


Figure 7. Percentage of correct exchanges with E1 and E2 by session for each subject. Only the first four exchanges of each trial in a session were counted, resulting in a total of 16 exchanges per session. Dotted line shows pass criterion (13/16 correct exchanges; 81.25% correct).

- ▲ Opaque panel used from this session onwards with E1
- ▲ Opaque panel used from this session onwards with E2

Although all the apes failed training, some of the individuals did surprisingly well despite not seemingly learning the conditional rule. This was likely due to some apes developing various strategies, as opposed to exchanging the objects randomly. For instance, Frederike and Alex seemed to develop a preferred object type; Frederike preferred the B type object in the first few sessions, successfully passing sessions two through four with E2, but then switched her preference to the A type object in sessions five through eight (see Figure 7B). Likewise, Alex developed a preference for the B type object from session six; however, this preference was lost from session nine, as his performance with E1 gradually improved (see Figure 7A). These performances likely reflected confusion by the apes as to which object was correct, but interestingly, they show that the apes did not resort to exchanging randomly; rather they chose one object type and persisted with it for a few sessions, and then switched and tried the other object type. Similarly, although Zira originally exchanged the objects randomly, in session nine she began to preferentially exchange the A object type and successfully passed a session with E1. The following sessions with E2 were her worst, whilst performance with E1 remained good, suggesting she developed a preference for exchanging the A type object (see Figure 7C).

The two most successful apes, Suaq and Padana, seemed to use a different strategy. These subjects were able to learn within a session which object type was rewarded. That is, in the first trial they would often exchange both object types randomly, then in the next three trials they would change their response to reflect the object that was currently being rewarded. This resulted in many passed sessions, but poor performance in the first few exchanges in which they learnt which object type was correct (see Table 4). This was also reflected by Suaq, who successfully passed seven sessions with E1, and six with E2, but only reached an average of 68.75% correct exchanges in his first four exchanges of his last four passed sessions. Despite failing to pass training, such performance shows an impressive ability to quickly update a response to reflect current reward contingencies.

Table 4

Number of Correct Exchanges in the First Four Exchanges of Each Trial, for Each Passed Session

	Session	Experimenter	Trial 1	Trial 2	Trial 3	Trial 4
Suaq	2	E1	2	3	4	4
	5	E1	4	4	4	4
	6	E1	2	4	3	4
	8	E1	3	4	4	3
	10	E1	1	4	4	4
	11	E1	3	4	4	4
	12	E1	4	4	4	4
	4	E2	3	4	4	4
	5	E2	4	4	4	3
	7	E2	2	4	4	4
	8	E2	4	4	4	4
	10	E2	1	4	4	4
	11	E2	4	4	4	4

Padana	2	E1	1	4	4	4
	4	E1	3	4	4	4
	6	E1	3	3	4	4
	7	E1	2	4	3	4
	8	E1	4	4	3	4
	9	E1	4	4	4	4
	11	E1	4	4	4	4
	1	E2	3	3	4	4
	2	E2	3	4	4	4
	8	E2	2	4	3	4
	11	E2	2	4	4	4
	12	E2	2	4	4	4

Note: Numbers in red indicate chance (or below chance) performance.

General Discussion

The purpose of the two experiments was to investigate whether apes can bind and recall social information. Using an object exchange paradigm, we taught apes to exchange a particular object type with a specific person. After a delay of two or ten weeks (Experiment 1), the apes' memory of which object type belonged to which person was assessed. Additionally, we tested their ability to use inference by exclusion by introducing a third person at test (Experiment 1), with whom they should give object type C; the type that did not belong to either of the other two experimenters.

The results from the two experiments combined show that the apes failed to bind and recall which object type was rewarded with which experimenter. In Experiment 1, the apes clearly remembered which objects were rewarded and could remember this information after long periods of time; however, they did not alter their response depending on which experimenter was present at test. The reason why they failed to bind the object types to the experimenters may have been due to learning a reversal discrimination rule during training, rather than discriminating between object types based on which experimenter was present. That is, they first learned that one object type was rewarded, and then learned that this object type was no longer rewarded, but instead one of the previously un-rewarded types was now rewarded. At test, they then preferentially exchanged the object that was rewarded last. However, the results from Experiment 2 suggest that the apes still failed to learn to exchange a particular object with a particular experimenter, even when there was no other obvious pattern or rule to guide their response during training.

Consequently, it seems that the apes simply did not readily encode or attend to which experimenter was present, and subsequently failed to bind this information to the object type. This failure to use social information may be specific to social information in the form of human experimenters. As previously mentioned, there is other evidence that apes do not incorporate the identity of experimenters as cues or information (Lewis et al., 2017a, b). For instance, Beran (personal communication) found that chimpanzees could not make use of social information to determine in which opaque container food was located. In one situation, chimpanzees failed to learn that one experimenter always carried an empty container, whilst the other experimenter always carried a container full of food. In another situation, they were unable to learn that one experimenter always pointed to the container with food inside, whereas the other experimenter always pointed to the empty container. Thus, the chimpanzees were unable to bind information about the identity of the experimenter to the location of food.

Nevertheless, there are cases in which chimpanzees and bonobos have learnt to successfully distinguish between the actions of different experimenters. For instance, Subiaul, Vonk, Okamoto-Barth, and Barth (2008) found that five of seven chimpanzees were able to learn that one individual would act

selfishly (by turning away from the ape and keeping food for themselves), and that another would act generously (by giving the food to the ape). Four of these chimpanzees were then able to successfully select a novel ‘generous’ experimenter after observing them acting generously (see also Russell, Call, & Dunbar, 2008). Similarly, Wobber et al. (2010, Experiment 3) found that chimpanzees and bonobos could successfully learn that one experimenter always held concealed food, and another did not. However, the success of the apes in these experiments may be a result of learning which actions/behaviors to respond to (e.g., learning to avoid any experimenter that turned away), rather than which specific person did what. Additionally, in the Wobber et al. (2010) study, the experimenters always stood in the same location during testing, thus the apes may have used the spatial location of the experimenter, rather than the identity of the experimenter, to determine where the food was (i.e., food is always located with the left experimenter).

In our lab, social information in the form of a human experimenter may be ignored due to the apes being tested by many different experimenters (and sometimes multiple experimenters in one test), who usually are not an integral part of the task. In Experiment 1, if the apes failed to attend to, or encode, information about which experimenter was present, then it made sense for them to select object types based on which type was rewarded last. Importantly, the finding that they still failed to use the social information in Experiment 2, despite having no other information to guide their choice, suggests that such information is not readily attended to or encoded. Consequently, it may be that the apes were able to bind social information, but did not use the opportunity to do so because they did not attend to, or regard, the social information as integral to the task. Indeed, Martinez and Matsuzawa (2009) found that chimpanzees were able to perform above chance in a conditional discrimination task when the social stimuli were a male human and a female chimpanzee, suggesting that apes may be able to use social stimuli successfully when such stimuli are more salient. However, the present finding that apes seem unable to readily encode the identity of humans, at least when the information is not salient, warrants further research using other task designs and ape populations to clarify the robustness of these findings. For example, it is possible that with more or different training, the apes may have learned that the experimenter was crucial to the task.

Although the apes failed to use the experimenter identity in our task, they did not resort to exchanging objects randomly. Instead, they performed in a consistent manner, choosing to exchange an object type that was rewarded most recently (Experiment 1), or by developing individual strategies that varied in success (Experiment 2; see previous discussion). Such performance was unexpected, but was logical and sensible if the identity of the experimenter was not encoded or attended to. These findings can be said to be “meaningful failures,” a term coined by Breland and Breland (1961), in which nonhuman animals respond in unexpected but meaningful ways.

A potential future avenue would be to conduct a similar test with conspecifics, due to previous research suggesting that conspecifics may be more salient as cues (Martinez & Matsuzawa, 2009). Indeed, early research concerning theory of mind suggested that apes were unable to adjust their behavior based on what a human experimenter could, or could not, see (Povinelli & Eddy, 1996), but were successful with conspecifics (Hare, Call, Agnetta, & Tomasello, 2000). Alternatively, the apes may learn to attend to the identity of the experimenter through scaffolding; as the apes’ previous testing experiences have likely predisposed them to ignore who it is that is testing them, and to fail to consider the experimenter as a stimulus in a task, they may require additional support to learn to encode this information¹. For instance, in this task, the experimenters could initially be made salient in some way (such as by wearing unusual clothing), so that the apes pay attention to who is exchanging objects with them. The saliency of the experimenters could then be reduced once the apes understand that the experimenter is an important element of the task. Furthermore, future research should examine whether the apes’ performance might be improved by training with additional experimenters and objects (i.e., more than two experimenters and two objects). This would make learning a reversal rule or updating within a session more complex (as there would be multiple alternative choices), and as such, may prompt

¹ We wish to thank an anonymous reviewer for this suggestion.

the apes to look for alternative information (such as the identity of the experimenter). Likewise, the apes may have benefited from training with both experimenters in the same session, that is, having E1 present in one area of the testing room, and E2 in another area (counterbalanced between sessions). This would require the apes to adapt their response within a session, as opposed to learning that one object type is consistently rewarded in a session, and again may result in the apes searching for additional information to determine which object is correct. Finally, it may be worthwhile to compare children's performance on a similar task, as there is evidence that humans may rely more on social information than apes (van Leeuwen, Call, & Haun, 2014).

In summary, we found that apes did not bind and recall which object type belonged to which experimenter, and consequently did not infer by exclusion that an untrained object type belonged to a new experimenter. This failure to bind and recall social information was not a result of forgetting, as the apes clearly showed a preference for the previously rewarded objects types over the non-rewarded object type. Also, this effect was shown after both two and ten weeks, showing that apes can successfully remember information over long time periods. The apes thus seemed to show a "meaningful failure," whereby in the absence of any salient information, they based their responses on a reinforcement rule (Experiment 1), or used other strategies, such as updating within a trial (Experiment 2).

Ethical Approval

The study was ethically approved by an internal committee at the Max Planck Institute for Evolutionary Anthropology. Animal husbandry and research comply with the "EAZA Minimum Standards for the Accommodation and Care of Animals in Zoos and Aquaria", the "EEP Bonobo Husbandry Manual", the "WAZA Ethical Guidelines for the Conduct of Research on Animals by Zoos and Aquariums" and the "Guidelines for the Treatment of Animals in Behavioral Research and Teaching" of the Association for the Study of Animal Behavior (ASAB).

Acknowledgements

The authors wish to thank the staff and ape keepers at Leipzig Zoo, with special thanks to Hanna Petschauer and Marika Skowronski for their roles as experimenters in this study. The authors thank the Danish National Research Foundation [DNRF89] for funding.

References

- Anderson, J. R. (1996). Discrimination learning and multiple reversals in young adult and older monkeys *Macaca arctoides*. *The Quarterly Journal of Experimental Psychology: Section B*, *49*, 193–200.
- Anderson, J. R. (1998). Social stimuli and social rewards in primate learning and cognition. *Behavioural Processes*, *42*, 159–175.
- Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. L. Bjork & R. A. Bjork (Eds.), *Handbook of perception and cognition* (pp. 237–313). San Diego: Academic Press.
- Beran, M. J. (2004). Long-term retention of the differential values of Arabic numerals by chimpanzees (*Pan troglodytes*). *Animal Cognition*, *7*, 86–92.
- Beran, M. J., & Evans, T. A. (2009). Delay of gratification by chimpanzees (*Pan troglodytes*) in working and waiting situations. *Behavioural Processes*, *80*, 177–181.
- Beran, M. J., & Heimbauer, L. A. (2015). A longitudinal assessment of vocabulary retention in symbol-competent chimpanzees (*Pan troglodytes*). *PLoS One*, *10*, e0118408.
- Beran, M. J., & Washburn, D. A. (2002). Chimpanzee responding during matching to sample: Control by exclusion. *Journal of the Experimental Analysis of Behavior*, *78*, 497–508.
- Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681–684.
- Cheney, D., Seyfarth, R., & Smuts, B. (1986). Social relationships and social cognition in nonhuman primates. *Science*, *234*, 1361–1366.
- Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*,

- 272–274.
- Clayton, N. S., Yu, K. S., & Dickinson, A. (2001). Scrub jays (*Aphelocoma coerulescens*) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology Animal Behavior Processes*, *27*, 17–29.
- Crystal, J. D., & Smith, A. E. (2014). Binding of episodic memories in the rat. *Current Biology*, *24*, 2957–2961.
- Eacott, M. J., Easton, A., & Zinkivskay, A. (2005). Recollection in an episodic-like memory task in the rat. *Learning & Memory*, *12*, 221–223.
- Fagot, J., & Cook, R. G. (2006). Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 17564–17567.
- Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*, 771–785.
- Itoh, K., Izumi, A., & Kojima, S. (2001). Object discrimination learning in aged Japanese monkeys. *Behavioral Neuroscience*, *115*, 259–270.
- Janmaat, K. R. L., Ban, S. D., & Boesch, C. (2013). Chimpanzees use long-term spatial memory to monitor large fruit trees and remember feeding experiences across seasons. *Animal Behaviour*, *86*, 1183–1205.
- Kano, F., & Hirata, S. (2015). Great apes make anticipatory looks based on long-term memory of single events. *Current Biology*, *25*, 2513–2517.
- Laska, M., Alicke, T., & Hudson, R. (1996). A study of long-term odor memory in squirrel monkeys (*Saimiri sciureus*). *Journal of Comparative Psychology*, *110*, 125–130.
- Lewis, A., Call, J., & Berntsen, D. (2017a). Distinctiveness enhances long-term event memory in non-human primates, irrespective of reinforcement. *American Journal of Primatology*, *79*, e22665.
- Lewis, A., Call, J., & Berntsen, D. (2017b). Non-goal-directed recall of specific events in apes after long delays. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170518.
- MacDonald, S. E. (1994). Gorillas' (*Gorilla gorilla gorilla*) spatial memory in a foraging task. *Journal of Comparative Psychology*, *108*, 107–112.
- Martin-Malivel, J., & Okada, K. (2007). Human and chimpanzee face recognition in chimpanzees (*Pan troglodytes*): Role of exposure and impact on categorical perception. *Behavioral Neuroscience*, *121*, 1145–1155.
- Martin-Ordas, G., Berntsen, D., & Call, J. (2013). Memory for distant past events in chimpanzees and orangutans. *Current Biology*, *23*, 1438–1441.
- Martinez, L., & Matsuzawa, T. (2009). Visual and auditory conditional position discrimination in chimpanzees (*Pan troglodytes*). *Behavioural Processes*, *82*, 90–94.
- Mendes, N., & Call, J. (2014). Chimpanzees form long-term memories for food locations after limited exposure. *American Journal of Primatology*, *76*, 485–495.
- Menzel, C. R. (1999). Unprompted recall and reporting of hidden objects by a chimpanzee (*Pan troglodytes*) after extended delays. *Journal of Comparative Psychology*, *113*, 426–434.
- Mineka, S., Davidson, M., Cook, M., & Keir, R. (1984). Observational conditioning of snake fear in rhesus monkeys. *Journal of Abnormal Psychology*, *93*, 355–372.
- Murai, C., Tanaka, M., Tomonaga, M., & Sakagami, M. (2011). Long-term visual recognition of familiar persons, peers, and places by young monkeys (*Macaca fuscata*). *Developmental Psychobiology*, *53*, 732–737.
- Normand, E., Ban, S. D., & Boesch, C. (2009). Forest chimpanzees (*Pan troglodytes verus*) remember the location of numerous fruit trees. *Animal Cognition*, *12*, 797–807.
- Noser, R., & Byrne, R. W. (2007). Mental maps in chacma baboons (*Papio ursinus*): Using inter-group encounters as a natural experiment. *Animal Cognition*, *10*, 331–340.
- Osvath, M., & Persson, T. (2013). Great apes can defer exchange: A replication with different results suggesting future oriented behavior. *Frontiers in Psychology*, *4*, 698.
- Parr, L. A., Winslow, J. T., Hopkins, W. D., & de Waal, F. B. M. (2000). Recognizing facial cues: Individual discrimination by chimpanzees (*Pan troglodytes*) and rhesus monkeys (*Macaca mulatta*). *Journal of Comparative Psychology*, *114*, 47–60.
- Pele, M., Dufour, V., Thierry, B., & Call, J. (2009). Token transfers among great apes (*Gorilla gorilla*, *Pongo pygmaeus*, *Pan paniscus*, and *Pan troglodytes*): Species differences, gestural requests, and reciprocal exchange. *Journal of Comparative Psychology*, *123*, 375–384.
- Pokorny, J. J., & de Waal, F. B. (2009a). Face recognition in capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology*, *123*, 151–160.
- Pokorny, J. J., & de Waal, F. B. (2009b). Monkeys recognize the faces of group mates in photographs. *Proceedings of the National Academy of Sciences, USA*, *106*, 21539–21543.

- Povinelli, D. J., & Eddy, T. J. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development*, *61*, 1–152.
- Rosenfeld, S. A., & Van Hoesen, G. W. (1979). Face recognition in the rhesus monkey. *Neuropsychologia*, *17*, 503–509.
- Russell, Y. I., Call, J., & Dunbar, R. I. (2008). Image scoring in great apes. *Behavioural Processes*, *78*, 108–111.
- Schwartz, B. L., Colon, M. R., Sanchez, I. C., Rodriguez, I. A., & Evans, S. (2002). Single-trial learning of “what” and “who” information in a gorilla (*Gorilla gorilla gorilla*): Implications for episodic memory. *Animal Cognition*, *5*, 85–90.
- Schwartz, B. L., Meissner, C. A., Hoffman, M., Evans, S., & Frazier, L. D. (2004). Event memory and misinformation effects in a gorilla (*Gorilla gorilla gorilla*). *Animal Cognition*, *7*, 93–100.
- Sliwa, J., Duhamel, J. R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 1735–1740.
- Smith, S. M. (1994). Theoretical principles of context-dependent memory. *Theoretical aspects of memory*, *2*, 168–195.
- Smith, S. M. (2013). Effects of environmental context on human memory. In T. J. Perfect & D. S. Lindsay (Eds.), *Handbook of applied memory* (pp. 162–182). London: SAGE Publications.
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220.
- Subiaul, F., Vonk, J., Okamoto-Barth, S., & Barth, J. (2008). Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition*, *11*, 611–623.
- Tomonaga, M. (1999). Inversion effect of perception of human faces in a chimpanzee. *Primates*, *40*, 417–438.
- Uher, J., & Call, J. (2008). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, *Gorilla gorilla*) perform on the reversed reward contingency task II: Transfer to new quantities, long-term retention, and the impact of quantity ratios. *Journal of Comparative Psychology*, *122*, 204–212.
- Vale, G. L., Flynn, E. G., Pender, L., Price, E., Whiten, A., Lambeth, S. P., ...Kendal, R. L. (2016). Robust retention and transfer of tool construction techniques in chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology*, *130*, 24–35.
- van Leeuwen, E. J., Call, J., & Haun, D. B. (2014). Human children rely more on social information than chimpanzees do. *Biology Letters*, *10*, 20140487.
- Vlamings, P. H., Hare, B., & Call, J. (2010). Reaching around barriers: The performance of the great apes and 3-5-year-old children. *Animal Cognition*, *13*, 273–285.
- Vlamings, P. H., Uher, J., & Call, J. (2006). How the great apes (*Pan troglodytes*, *Pongo pygmaeus*, *Pan paniscus*, and *Gorilla gorilla*) perform on the reversed contingency task: The effects of food quantity and food visibility. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 60–70.
- Völter, C. J., & Call, J. (2017). Causal and inferential reasoning in animals. In J. Call, G. M. Burghardt, I. M. Pepperberg, C. T. Snowdon, & T. Zentall (Eds.), *APA handbook of comparative psychology Vol 2: Perception, learning, and cognition* (pp. 643–671). Washington, DC: American Psychological Association.
- Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 1–15.
- Wobber, V., Wrangham, R., & Hare, B. (2010). Bonobos exhibit delayed development of social behavior and cognition relative to chimpanzees. *Current Biology*, *20*, 226–230.